VOLUME 22 NUMBER 2 JUNE 1959

SOCIOMETRY

A JOURNAL OF RESEARCH IN SOCIAL PSYCHOLOGY

Articles by

DEUTSCH and SOLOMON
ROSENGREN
ROSENBERG
ROSS and HARARY
BORGATTA
CHAPMAN and CAMPBELL
PHILP and DUNPHY
MIDDLETON

Published by

THE AMERICAN SOCIOLOGICAL SOCIETY

SOCIOMETRY

EDITOR

LEONARD S. COTTRELL, JR.

EDITOR-ELECT (July 1959)

JOHN A. CLAUSEN

ASSOCIATE EDITORS

EDGAR F. BORGATTA
LAUNOR F. CARTER
LEON FESTINGER
NELSON N. FOOTE
NATHANIEL L. GAGE
ERVING GOFFMAN
FREDERICK MOSTELLER
HERBERT A. SIMON
FRED L. STRODTBECK
GUY E. SWANSON

EDITORIAL CONSULTANTS

WENDELL BELL
PETER M. BLAU
ORVILLE G. BRIM, JR.
ROGER W. BROWN
JAMES S. COLEMAN
SANFORD M. DORNBUSCH
FRED E. FIEDLER
ALLAN R. HOLMBERG
ALEX INKELES
MELVIN L. KOHN
WILLIAM W. LAMBERT
DUNCAN MACRAE, JR.
JAMES G. MARCH
THEODORE R. SARBIN
STANLEY SCHACHTER
WILLIAM A. SCOTT
SIDNEY SIEGEL
SHIRLEY A. STAR
RALPH H. TURNER
EUGENE A. WEINSTEIN
ROBIN M. WILLIAMS, JR.
ROBERT F. WINCH
MARIAN RADKE YARROW

EXECUTIVE OFFICER

MATILDA WHITE RILEY

Published four times a year

March, June, September, December

by

THE AMERICAN SOCIOLOGICAL SOCIETY

at The Boyd Printing Co., Inc., 49 Sheridan Ave., Albany 10, N. Y.

Sociometry was founded in 1937 by J. L. Moreno

Communications for the Editors and all manuscripts should be addressed to John A. Clausen, The National Institutes of Health, The Clinical Center, Room 2N208, Bethesda 14, Maryland. Business communications should be addressed to Matilda White Riley, The American Sociological Society, New York University, Washington Square, New York 3, New York.

Copyright © 1959 by THE AMERICAN SOCIOLOGICAL SOCIETY

Second class postage paid at Albany, N. Y.

SOCIOMETRY

A JOURNAL OF RESEARCH IN SOCIAL PSYCHOLOGY

VOLUME 22 · NUMBER 2 · JUNE 1959

CONTENTS

Reactions to Evaluations by Others as Influenced by Self-Evaluations
Morton Deutsch and Leonard Solomon 93

Symptom Manifestations as a Function of Situational Press: A Demonstration in Socialization ..... *William R. Rosengren* 113

The Maintenance of a Learned Response in Controlled Interpersonal Conditions ..... *Seymour Rosenberg* 124

A Description of Strengthening and Weakening Members of a Group
Ian C. Ross and Frank Harary 139

A Broad Empirical Approach to Personality: *Personality and Motivation Structure and Measurement* ..... *Edgar F. Borgatta* 148

The Effect of Acquiescence Response-Set upon Relationships among the F Scale, Ethnocentrism, and Intelligence
Loren J. Chapman and Donald T. Campbell 153

Developmental Trends in Small Groups
Hugh Philp and Dexter Dunphy 162

Negro and White Reactions to Racial Humor ..... *Russell Middleton* 175

SOCIOMETRY

EDITORIAL POLICY

Sociometry is concerned with the entire range of interests and problems represented by research in social psychology. It is the policy of the editors to seek those manuscripts for publication which represent the significant research interests of investigators who are concerned with giving the field of social psychology theoretical structure and reporting research which is clearly focused, well designed, and competently conducted.

While social psychology is presently regarded by most as a field with indeterminate boundaries, it has as its central focus the investigation of the processes and products of social interaction at the interpersonal, intrapersonal, intergroup, and intragroup levels and the development of significant generalizations therefrom. In keeping with the more general meaning of the name of the journal emphasis will be placed on measurement of social behavior. However, this emphasis does not exclude the acceptability of good articles which must rely upon qualitative materials and analyses.

The editors and editorial consultants can be expected to subject manuscripts to rigorous criticism and screening according to the best standards of scientific research and at the same time avoid a sterile orthodoxy which would stultify the communication of creative ideas at the growing edge of the science. Thus the journal will strive to be flexible in its response to the publication needs of its contributors.

It is the intention of the editors to avoid any tendency toward professional provincialism and to invite contributions from any sector of the scientific community which promise to further the objectives of the journal.

Annual *subscription* rate $9.00; special rate to members of THE AMERICAN SOCIOLOGICAL SOCIETY $4.50. Single issues $2.50.

Specifications for Manuscript Preparation for Sociometry may be obtained from the Editorial Office.

Back Issues dated 1937 through 1955 should be ordered from the former publisher, Beacon House, Beacon, New York.

Notice of *change of subscriber's address* should be sent to THE SOCIETY four weeks in advance, indicating the old as well as the new address.

Reactions to Evaluations by Others as Influenced by Self-Evaluations

MORTON DEUTSCH, *Bell Telephone Laboratories*
LEONARD SOLOMON,[1] *Bell Telephone Laboratories*

It has long been recognized that our perceptions and evaluations of other people may be profoundly influenced by our perceptions and evaluations of ourselves. Recently, psychologists have turned their attention to this phenomenon. The studies have largely focused upon the *distortions* which are introduced into interpersonal perception by such characteristics of the perceiver as his emotional state, his needs, his expectations, his attitudes, his social role, his lack of self-insight, etc.

To a certain extent, the emphasis upon the problem of accuracy in interpersonal perception, unwittingly, has tended to foster the view that the characteristics of the perceiver, e.g., his needs and values, enter into the perceptual process primarily as distorting influences. It seems more reasonable to assume that both veridical and nonveridical perceptions are the outcomes of the individual's attempt to achieve a percept which is consistent with the various influences operating upon him, i.e., influences from the stimuli emanating from the object to be perceived and from the object's context, influences from the residues within the perceiver of his past experiences, and influences from the perceiver's psychological state.

The significance of the need for consistency or balance has been noted by many writers who have often used different terms for the same concept. Heider (4) uses the phrase "cognitive balance," Festinger (3) expresses it as "consonance," Osgood, Suci, and Tannenbaum (11) state it as "congruence," Lecky (9) employs "self-consistency," still others have used such terms as "harmony," "equilibrium," "stability," "integration" to express similar notions. The essential idea underlying these various terms is that the individual attempts to perceive (cognize, or evaluate) the various aspects of his environment and of himself in such a way that the behavioral implications of his perceptions are not in contradiction. The existence of such contradictory implications produces a state of imbalance ("dissonance," "incongruity," "disharmony," etc.).

Heider (4) has indicated that whenever a state of imbalance exists or impends, a tendency will be produced in the individual to engage in behavior

[1] Dr. Solomon is now on the staff at the V. A. Hospital at Montrose. We wish to acknowledge gratefully the assistance of Mrs. Ann Griffith and Mr. George Rotter in the analysis of the data.

which will change the perceived entities (or the relations among the entities) in such a way as to remove or prevent imbalance. Under conditions which are perceived as not permitting or not making feasible such behavior, a tendency will be produced in the individual to change his perceptions directly.

To illustrate, let us suppose that one hears from an acquaintance that a good friend has stolen money from a blind beggar. This information is inconsistent with the deepseated beliefs one holds concerning one's friend. Various possibilities of preventing or removing the inconsistency exist. A primitive method of dealing with the inconsistency would be to "not hear" or to misunderstand what is said. If one "hears" then one might attempt to convince one's acquaintance that he is wrong, that it was somebody else and not one's friend who did it. If he is adamant and refuses to retract his statement, one may attempt to discredit the acquaintance or his statement, e.g., one may think he did not really observe what he reported or that he dislikes one's friend and is lying. One may even seek out new sources of information which will help to discredit the negative statement about one's friend. If the statement cannot be discredited, one may accept it with the qualification that it does not have the implication such an action usually has. One may admit that one's friend did steal the money but think that it was a "joke" and that he returned the money, or one might think that he was forced to do it at the point of a gun and that he felt guilty about it, or one might think that he was not himself when he did it—he was drunk. If the constraints of reality are such that one is unable to avoid, distort, negate, or reinterpret the negative information about one's friend one may be forced to change one's views of one's friend. If this occurs, one may then attempt to obtain new information which will help to support the newly acquired negative conception.

Imbalance or inconsistency produces a tendency to restore balance. Something must yield if balance is to be restored; if not, the individual will be in a state of tension. But what will yield? Various theorists have attempted to formulate answers to this question. Krech and Crutchfield (8) have stated that "other things being equal, a change introduced into the psychological field will be absorbed in such a way as to produce the smallest effect on a strong structure." Osgood *et al.* (11) suggest that the elements in conflict may all yield and state that the magnitude of the shift will be inversely proportional to the intensities of the interacting units. Festinger (3) has written that "the maximum dissonance that can possibly exist between any two elements is equal to the total resistance to change of the less resistant element." He indicates that the resistance of an element to change is a function of the number and importance of the elements which are consonant with it. Similarly, Deutsch (2) has stated: "When two cognitions are in contradic-

tion to one another, the cognition with the weaker resistance to change will change. The resistance to changing a cognition will be directly related to the number and strength of other cognitions with which it is consonant and to the strength of the anticipated positive motivational consequences based upon it and inversely related to the strength of the anticipated negative motivational consequences of behavior based upon it."

None of the foregoing answers to the question of "what will yield" are sufficiently specified to enable precise predictions. However, they suggest that deeply rooted perceptions are likely to be the focal points around which other, interrelated, perceptions get organized. That is, there will be a tendency to form new perceptions in such a way that they are not inconsistent with existing, strongly rooted structures within the individual.

It is the thesis of this paper that perceptions, cognitions, and evaluations of aspects of the self are, frequently, the focal points around which other relevant perceptions get organized. This thesis is implicit in the use of projective tests as diagnostic tools; the clinician by observing how the individual organizes ambiguous perceptual material is able to make inferences to the strongly anchored self-perceptions which serve as focal points for the organization of the ambiguous material.

In this paper, we are concerned with *self-evaluation* and the influence of self-evaluation upon the evaluation of others. Our study originates with the view that one can only predict how an individual will react to another's evaluation of some aspects of him *if* one knows how the individual evaluates this aspect himself. Many social psychologists assume implicitly in their statements of the various implications to be drawn from their theories that the individual evaluates himself positively; rather different predictions would be made from the opposite assumption.

Thus, one of Heider's fundamental propositions is stated as follows (4): "a balanced state will exist if sentiment and belongingness go together, that is, if p is united with the persons he likes and if he likes the persons he is united with." The hidden assumption in this proposition is that p *likes* himself (i.e., evaluates himself positively).[2] If p *dislikes* himself (i.e., evaluates himself negatively), one would expect just the opposite: namely, that a balanced state would occur if he is united with the persons he dislikes and if he dislikes the persons he is united with. Similarly, from Heider's formulations one would expect a tendency in an individual (p) to like another (o) if p perceives that

[2] In Heider's most recent work (5), published after this study was conducted, he explicitly states that most of his examples presuppose a positive attitude toward the self. He writes (p. 290): "However the possibility of a negative attitude toward the self . . . must also be considered. One would expect it to play a role contrary to that of [the person liking himself]."

both he and o like the same thing (e.g., another person, a style of painting, a philosophical viewpoint). Here again it is assumed that p likes himself. If p dislikes himself, one would expect just the opposite: a tendency to dislike or to be contemptuous of o, if o likes what p likes.

In drawing implications from his theory of dissonance, Festinger also implicitly assumes that the individual likes himself. Thus, he indicates that the person who continues to smoke, knowing that it is bad for his health, will be in a state of dissonance unless he can somehow rationalize his continued smoking. This statement, of course, assumes that the person likes himself; if he dislikes himself, there would be no inconsistency between his continuing to smoke and knowing that it is bad for his health. Similarly, the results of Brehm's experiment as cited by Festinger (3) are in line with the theory of dissonance only if one assumes that the subjects liked themselves. Brehm's findings indicate that people, after making a choice between two valued objects, tend to devalue the unchosen object and enhance the value of the chosen object. If the subjects disliked themselves, one would expect the opposite result: namely, that they would depreciate the values of the object which was chosen and enhance the value of the unchosen object. Undoubtedly, it is more common that people choose what they like and like what they choose. However, work with neurotic patients, who dislike aspects of themselves, would suggest that it is possible for people to choose what they dislike and to dislike what they choose.

The assumption that the individual likes himself is also implicit in one of Homans' basic propositions (6): "If the frequency of interaction between two or more persons increases, the degree of their liking for one another will increase, and vice versa." If an individual dislikes himself, one would expect the opposite: the more interaction there is between himself and another, the more his degree of contempt or disliking will increase. Similarly, reinforcement theories often implicitly assume that social approval and praise serve as rewards which reinforce stimulus-response connections independently of the individual's self-evaluation. The results of our experiment will suggest that the significance to the individual of such rewards as praise may be opposite if the individual's self-evaluation is negative rather than positive.

Many phenomena of everyday life which, otherwise, appear to be puzzling seem more understandable if one assumes that the person involved evaluates himself or herself negatively rather than positively.[3] Thus, it is difficult ordinarily to understand why a girl may reject men who are attracted to her

[3] It may be noted that such theories as those offered by Heider, Festinger, Homans, etc., might help to explain a wide range of puzzling and pathological phenomena if the assumption of negative self-evaluation were substituted for the assumption of positive self-evaluation.

and be attracted to men who reject her, or why a child may seek punishment, or why a man may feel a sense of comfort when he has been defeated in a competition or failed an important test and feel uncomfortable when he has been victorious or successful. However, if we assume that the girl, the child, or the man who act in such puzzling ways evaluate themselves negatively the bewildering quality of their actions disappears; their actions seem comprehensible. This is not to assert that such phenomena are completely explained by the assumption of a negative rather than a positive self-evaluation. Other factors, particularly the indirect expression of hostility toward others, may also be involved. Nevertheless, it seems likely that many behavioral phenomena which seem incomprehensible or "irrational" have at their roots a self-evaluation by the behaving individual which is counter to that which the observer assumes him to have. Some unpublished studies (2) indicate that untrained observers find it difficult to arrive at the conception that a person they are observing has a negative evaluation of himself even when the evidence is strong.

So far, we have indicated the importance of deeply rooted perceptions, cognitions, or evaluations of self as a focal point which may influence how we perceive, cognize, or evaluate the environment with which we interact. We have expressed the view that there is a need to perceive the various aspects of one's environment and of oneself in such a way that the behavioral implications of one's perceptions are not in contradiction. From the need for consistency, with the assumption that a self-evaluation is deeply rooted, one would hypothesize that an individual will tend to evaluate another (who reacts to him positively or negatively) in such a way that his evaluation of the other is consistent with or appropriate to his own self-evaluation. More specifically, we would hypothesize that *if an individual evaluates some aspect of himself (as positive or negative) and another evaluates it similarly, the individual will tend to evaluate the other person favorably; if their evaluations are dissimilar, the individual will tend to evaluate the other unfavorably.* Hence, one would expect that if a person evaluates negatively some aspect of himself he will tend to esteem those who regard this aspect of him negatively; such a person, on the other hand, would tend to view negatively another who esteems an aspect of himself he dislikes. In a complementary way, one would expect that an individual who esteems some aspect of himself will tend to regard favorably another whose evaluation is also positive and to consider unfavorably another whose evaluation is negative.

The experiment, described below, was designed to test the hypothesis stated in the preceding paragraph. To test the hypothesis, it is necessary to create, experimentally, the circumstance in which an individual is evaluating *some aspect of himself*. We shall not digress into an examination of the concept of

"self" except to state that the self of an individual is composed of the symbolic representations he has made to himself of his various characteristics such as his physical, biological, psychological, ethical, and social characteristics. The symbolic representations the individual has made usually include representations of things with which he identifies or is identified with, such as his actions, productions, memberships, possessions, etc. Since the work of Cooley (1) and Mead (10), it is generally assumed that the establishment of the particular qualities of an individual's self, i.e., what is selected for self-representation and the nature of the representation of what is selected, is largely determined by the reactions of others upon whom the individual is dependent. The preceding several sentences provide the basis for assuming that an individual's evaluation of his performance on a task which is being socially evaluated provides a reasonable experimental analogue of "evaluating some aspect of himself."

One further point. The methodology of this study is experimental. An alternative procedure, of course, would have been to select subjects so that they differed systematically in their evaluations of aspects of themselves. The difficulties of testing a hypothesis unambiguously in nonexperimental designs of this kind are discussed fully in Selltiz, Jahoda, Deutsch, and Cook (13).

We have, in effect, attempted to create as independent variables (i.e., through the experimenter's manipulations) the condition in which a subject either has a favorable or unfavorable evaluation of some aspect of himself. Does our preceding discussion suggest that this is impossible, that a person's "general" evaluation of himself will influence his evaluation of any specific aspect of himself? Prior to the experiment, we were uncertain as to how this question should be answered so we obtained a measure of the subject's general self-esteem before he was exposed to the experimental variables. A priori, there seem to be several major possibilities with regard to the effects of the individual's self-esteem upon his evaluation of his performance on a given task. One possibility is that the experimental situation is sufficiently encapsulated from the other life situations of the subject to be relatively immune from his enduring personality characteristics. Another possibility is that initial self-esteem establishes the subject's level of expectation which, in turn, determines his evaluation of his performance. If he did better than he expected, his evaluation would be positive and if he did worse his evaluation would be negative. Since our experimental procedure entailed giving the subjects in the "success" treatment extremely good scores while the subjects in the "failure" treatment received extremely poor scores in comparison with their peer group, it would seem likely that the experimental treatments would be effective for most subjects in establishing the appropriate direction (positive or negative) of the evaluation of their performances.

A further possibility is that the subject's evaluation of his own performance will tend to be made consonant with his initial self-esteem. If this tendency were operating and if the subject's perception of his own performance were not strongly anchored, one would expect the experimental treatments to have little effect. However, this expectation is not appropriate if the subject's perception of his performance is strongly anchored. When the subject's self-esteem is inconsistent with his perception of his performance and they both are strongly anchored, he would be in conflict and could be expected to dissociate himself from his performance, attributing it to factors outside of himself such as "chance" or "accident." (Were it not for some such mechanisms, the individual's self-esteem would be rather unstable.) Initial self-esteem then becomes irrelevant since the performance is not connected with the self. On the other hand, if the subject's performance is consistent with his view of himself, instead of dissociating himself from it one would expect him to identify with it. Hence, either because he dissociates himself from performances which are inconsistent or identifies himself with performances which are consistent with his view of himself, one would expect the experimental treatments to give the results predicted from our hypothesis.

METHOD

Subjects

All Ss were employees of the Long Lines Department of the American Telephone and Telegraph Company. The Ss were assigned to participate in the experiment during their regular working hours. They were free not to participate but practically no Ss refused to participate. The main experiment was conducted with 132 women who were employed as telephone operators; they ranged from seventeen to over fifty years in age and had been employed from a few months to over 30 years. The experiment was replicated with 70 male Ss, employed as craftsmen, who ranged in age from twenty-three to thirty-four years. They had been employed from 2 to 14 years. For brevity's sake, the results of the replication will not be presented since its results are essentially the same as the main experiment.

Procedure

Eight Ss at a time came to an experimental room. The eight Ss were of the same sex and were relatively homogeneous with respect to age and experience. Each S was seated behind a desk surrounded by partitions in such a way that she could not see any other S. The Ss were then assured that the experiment would not affect their position in any way, that they would not be asked to write their names or any other identifying data on any of the questionnaires,

etc. The Ss were then asked to fill out the Tennessee Department of Mental Health's *Self-Concept Scale* (12). This scale is composed of 100 items. From the S's rating of the truth or falsity of each item as it applies to himself, it obtains measures of "what he is," "how he feels about himself," and "what he does" with regard to five areas: his physical characteristics, his moral-ethical characteristics, his psychological characteristics, his primary group membership, and his secondary group memberships. It also obtains a measure of the Ss candidness in filling out the questionnaire. The scale was used in an attempt to obtain a measure of the Ss general self-evaluation.

After the administration of the *Self-Concept Scale,* the Ss were told that the purpose of the experiment was to study how people form impressions of each other, especially when they have only very limited information to go on. The purpose of the partitions, it was explained, was to prevent them from forming impressions of each other on any other basis than from the way the other people work and from the written notes they would get from others. The Ss were then told that:

"You will be formed into two teams, Team A and Team B, that will compete with each other. Four of you will be on Team A, four on Team B. There will be no prize for the winning team but I want you to imagine that there is some important prize—such that each of you on the winning team will get $10. The winning team will be the one that wins two out of three contests. In each contest you will be given some problems to do and we want to see how many problems you can do correctly in a given amount of time. Each person will be given the same problems to do. The final score of each team for each contest will be the sum of the scores for each member of the team. Thus, if any person does very well she may bring her team ahead of the other team and help her team win, or if a person does very poorly, she may make her team fall behind the other team. At the end of each contest I will let you know which team has won."

Then, by a note handed to them, the Ss were informed which team they would be on and what number on the team they would have. It was explained that they would not be able to tell by the seating arrangements what the team or position number was of any other S. All Ss were led to believe they were on Team A in Position 4.

The Ss were told that in the first contest each one would be given a task that would measure the ability to do flexible thinking. The significance of this ability in helping people avoid getting into ruts was stressed. The *Gottschaldt Embedded Figures Test* was used. In this test, Ss have to find, for each of 39 problems, a simple geometrical figure which is embedded in a more complicated figure. They were told that the scores were weighted scores, depending upon the difficulty of the item, and hence the score they obtained could be

higher or lower than the number done correctly. Ss were given 8 minutes for the test, after which their test booklets were collected and "scored." They were then individually handed "scorecards" which contained "scores" for each of the positions on Team A and Team B as well as totals for each team. Each S was given, as her own score, a score which approximated the number of problems she had done correctly. Since each S in the room had been led to believe she was No. 4 on Team A, the seven other scores (all four of Team B and three of Team A) were made up by E to create the following experimental treatments:

A. *Individual Success, Group Success (IS, GS).* S was led to believe that she obtained the highest score among the eight people on both teams and that her team won the contest as a result. The made-up scores were arranged so that the S's own score was slightly less than twice as high as the average score of all the others; apart from her score the average scores for Team A and Team B were approximately the same.

B. *Individual Failure, Group Failure (IF, GF).* S was led to believe that she obtained the lowest score among the eight people on both teams and that her team lost the contest as a result. S's score was slightly less than half as high as the average score made up for the others; apart from S's score, the averages of Team A and Team B were approximately the same.

C. *Individual Success, Group Failure (IS, GF).* S was led to believe that she got the highest score but, even so, her team lost. S's score was slightly less than twice the average of the seven other scores; the differences between Team A and Team B were arranged so that Team B beat Team A by one or two points despite S's score.

D. *Individual Failure, Group Success (IF, GS).* S was led to believe that she got the lowest score but, even so, her team won. S's score was slightly less than half of the average score of the seven others; the differences between Team A and Team B were arranged so that Team A beat Team B by one or two points despite S's score.

When the Ss were given their scorecards, E announced which team had won the contest and pointed out that each S's score was encircled in red on her own scorecard. It was pointed out that they could tell from the scorecards how well each person had done.

After the Ss had some time to absorb the scorecards, they were administered the task for the second contest, the *Scrambled Words Test.* This test consisted of 40 words whose letters were rearranged (e.g., "after" was rearranged to read "teraf"); the task of the S was to arrange the scrambled letters so to form as many words as she could in 5 minutes. The Ss were told that this test was also a measure of flexibility of thinking and its significance for everyday functioning was stressed. After the test was finished, each S's

answer sheet was "scored" and scorecards were given to the Ss. The scorecards were made up to provide the S with the same kind of experience as she had in connection with the first task, i.e., if she were exposed to an IS, GS treatment on the first task, she also received an IS, GS treatment on the second task, etc.

After the Ss had read their scorecards and the contests were over (each S had been on a team that either won or lost the first two contests so there was no need for a third contest), the Ss filled out rating scales, indicating their evaluation of their own performance and their expectation of how they themselves and their performance would be evaluated by their teammates. Then the Ss were asked to write a note to some other person on their team.[4] They were told to express frankly their feelings about the performance of the person to whom they were to write and to indicate whether they wanted the note receiver on their team again if they were to have another contest. They were free to write anything else they wanted to in the note. What they wrote and how they wrote it would be a basis for the note receiver's forming an impression of their personality; similarly, what was said in the note they received would be a basis for forming an impression of the note-sender.

After the Ss wrote their notes, they were collected and the E substituted notes which he had prepared prior to the experiment. Half the Ss were given positive notes, half were given negative notes. Three different versions of the positive and negative notes were employed. The first version was used on about 10 per cent of the Ss, the second on 50 per cent, and the third on the remaining 40 per cent.[5] The *positive* notes follow:

1. I see that you have the highest (lowest) score for the team. I really think you did a fine job considering your situation. You were a valuable member of the team and I would very much like to have you on my team again.

2. You made a very high (low) score for our team in both contests and (but) I was pleased with your score. You probably will make a high (low) score again and (but) I certainly would like you on the same team with me again.

3. You are the person I most prefer to have on my team again.

The *negative* notes were:

1. I see that you got the high (low) score for the team but (and) I think you did a poor job considering the circumstances. I am very much opposed to having you on my team again.

[4] Some of the Ss were given instructions which specified to whom they were to write, other Ss were given instructions which allowed them to choose to whom to write. The instructions were distributed throughout the various experimental treatments. Since the different instructions produced no apparent differences, the differences in instructions will be ignored.

[5] The several versions of the notes were distributed throughout the various experimental treatments in a balanced way. The different versions produced no apparent differences in results and, hence, the three positive notes are grouped together as "positive" and the three negative notes are grouped as "negative."

2. You made a very high (low) score for our team on both contests but (and) I was displeased at your score. You will probably make a high (low) score again but (and) I would certainly not like to have you on the same team with me again.
3. You are the person I least prefer to have on the same team with me again.

After the Ss received the notes, they were instructed to form as strong an impression as they could of the note-writer and then they were asked to rate the note-writer on two types of rating scales. One set of scales were 9-point scales, ranging from "very high in this characteristic" to "very low in this characteristic," in which S was to rate the note-writer on the following characteristics: desirability as a teammate, intelligence, team spirit, desirability as a friend, and likeability. The other set of scales were in the form of the Semantic Differential (10). It consisted of 50 bipolar 7-point scales which included scales heavily loaded with 3 factors identified by Osgood, *et al.*: An *evaluative* or *goodness* factor ("good-bad," "nice-awful"), an *activity* factor ("fast-slow," "active-passive"), and a *strength* factor ("weak-strong," "timid-forceful"). On an a priori basis other scales were included to measure *warmth* ("warm-cold," "affectionate-unaffectionate"), *status* ("significant-insignificant," "important-unimportant"), *effectiveness* ("masterful-fumbling," "effective-ineffective"), *anxiety* ("afraid-unafraid," "worried-unworried") and *self-control* ("neat-sloppy," "conscientious-careless").

After the Ss had rated the note-sender on the Semantic Differential they were asked to rate, on a retrospective basis, how they felt about themselves when they learned what scores they and the others had obtained on the tests.

When they finished their self-rating the experiment was concluded and the E explained to the Ss what had taken place during the course of the experiment and discussed the purpose of the experiment and its everyday parallels. The cooperation of the Ss in not discussing the experiment was asked and apparently elicited. The Ss appeared to enjoy participating in the experiment.

To sum up the procedures in terms of the experimental design, a $2 \times 2 \times 2$ factorial design was created. The variables taken singly were: individual success (IS) versus individual failure (IF); group success (GS) versus group failure (GF); receiving a positive note (P) versus receiving a negative note (N). To facilitate computations, the number of cases in each cell were equalized by random elimination of cases in the cells with excessive numbers. After elimination, 14 women were left in each of the 8 cells of the factorial experiment; 6 men were left in each of the 8 cells of the male factorial design.

RESULTS

Self-evaluation

It is appropriate to examine at the outset the data that are relevant to the Ss' evaluations of their performances to determine whether or not our

experimental manipulations produced the differences in self-evaluations necessary to test our basic hypothesis. From Table 1, it is apparent that the Ss who were exposed to the IS treatment evaluated their performances favorably, while the Ss in the IF treatment evaluated their performances unfavorably. The differences in self-evaluations are statistically significant ($p<.01$). Similar differences are also found with respect to the Ss' anticipated evaluations from their teammates in terms of their "desirability as a teammate," "intelligence," "team spirit," and "desirability as a friend"; in each case, the Ss in the IS groups expect more favorable evaluations than the Ss in the IF groups.

TABLE 1

S's Evaluation of Her Performance and the Evaluations She Anticipated from Her Teammates on Various Characteristics

	Experimental Treatment *				
Females (N = 112)	IS GS	IS GF	IF GS	IF GF	Statistically Significant Comparisons ‡ ($p < .05$)
1. Evaluation of own performance	4.9†	3.9	6.9	7.3	I; I/G
2. Desirability as a teammate	3.6	3.3	5.6	7.0	I; I/G
3. Intelligence	5.1	4.5	5.7	6.2	I; I/G
4. Team spirit	4.5	4.4	5.7	5.9	I
5. Desirability as a friend	3.8	3.6	4.3	4.4	I

* IS = Individual success; IF = Individual failure.
GS = Group success; GF = Group failure.
† All scales have their most favorable rating at "1" and their least favorable rating at "9."
‡ I = Individual performance; G = Group performance.

The first-order effects upon self-evaluation of being a member of a group that won rather than lost were comparatively negligible for most of the measures. The relative lack of significant results is consistent with the fact that the teams were artificial and permitted little basis for identification: the team members did not know each other or see each other during the experiment; they were interdependent only with respect to an imaginary prize. However, there was a significant interaction effect between individual and group performance such that the women evaluated their own performances (and also expected to be evaluated by others) most favorably if their team had failed and they had done well but least favorably if their team had failed and they had done poorly.[6]

[6] These latter results were not obtained with the male Ss. The men showed a slight, but not reliable, tendency to evaluate their performances more favorably when their teams won rather than lost.

Evaluation of the Note-writer

The results cited in the preceding section indicate that we succeeded in inducing the differences in self-evaluation which were necessary to test the hypothesis that individuals will tend to evaluate others in a way which is consistent with their self-evaluations. We obtained a variety of ratings of the note-writer; the results of these ratings are summarized in Table 2.

TABLE 2

S's Evaluation of the Note-writer on Various Characteristics as Influenced by the Different Experimental Treatments

Females (N = 112)	Experimental Treatment *				Statistically Significant Comparisons † ($p < .05$)
	P IS	P IF	N IS	N IF	
1. "Evaluative" factor	1.4	2.3	3.5	2.3	M; M/I
2. Like-dislike	2.0	3.2	4.7	2.8	M; M/I
3. Desirability as a friend	2.9	4.5	6.8	3.8	M; M/I
4. Desirability as a teammate	2.8	3.7	6.4	3.4	I; M; M/I
5. Team spirit	2.1	3.0	6.6	2.8	I; M; M/I
6. Intelligence	3.7	3.3	5.5	2.6	I; M/I
7. "Effectiveness" factor	2.5	2.4	3.6	2.3	I; M/I
8. "Activity" factor	2.7	1.9	3.6	2.1	I; M
9. "Status" factor	2.6	3.1	4.2	2.9	M; M/I
10. "Strength" factor	3.0	2.7	2.5	2.9	
11. "Warmth" factor	2.0	3.5	4.9	3.5	M; M/I
12. "Self-control" factor ‡	1.8	2.2	3.3	1.8	M; M/I

* P = Positive note; N = Negative note.
IS = Individual success; IF = Individual failure.

† M = Note received; I = Individual performance.

‡ The experimental treatment means for the "Self-control" factor are based on an n of 24; total N is 96.

NOTE: For scales 1, 4, 5, and 6 the most favorable rating is at "1" and the least favorable at "9"; for the other scales the most favorable rating is at "1" and the least favorable at "7."

First, let us examine the ratings of the note-writer which are most purely evaluative and, hence, which bear most directly upon our hypothesis. These are the "evaluative factor" (composed of the scales "good-bad," "nice-awful") and the ratings of "like-dislike" and "desirability as a friend."

The note-writer was evaluated most favorably in the P, IS treatment, somewhat less favorably in the N, IF and P, IF treatments (which differed only slightly in their ratings) and rather unfavorably in the N, IS treatment.[7]

[7] The results for the men were essentially the same except that the women tended to rate the note-writer more favorably in the N, IF than P, IF treatment where the reverse was true for the men.

The statistical analysis of the data indicates significant interaction effects between the nature of the note received and the experience of individual success or failure.[8] In addition, there is a statistical effect of the note per se such that the "positive note-writers" are evaluated more favorably than the "negative note-writers."

These data allow two alternative interpretations, one of which supports our hypothesis and one which qualifies the hypothesis considerably. The latter interpretation would be that the data indicates that the hypothesis is valid only when there is a positive self-evaluation and that when the S's self-evaluation is negative she will react in, essentially, the same way to a positive or negative evaluation from another. This interpretation would imply that the note per se has no independent effects, that its apparent effects are simply due to the interaction between the note and the individual success treatment. The obtained statistical differences between the positive and negative notes are in fact due to the statistical differences between the "P, IS" and "N, IS" treatments.

The alternative interpretation which supports our original hypothesis is that there is both the interaction effect as predicted by the hypothesis of cognitive balance and a "positivity" effect such that the Ss tended to evaluate the positive notes more favorably than the negative notes. Were the positivity effect operating by itself one would have expected the Ss who received a positive note (P, IS and P, IF) to have rated the note-writer more favorably than those who received negative notes (N, IS and N, IF); were the tendency for cognitive balance also operating by itself one would have expected the Ss who received notes which are consistent with their own evaluations (P, IS and N, IF) to have rated the note-writer more favorably than those who received inconsistent notes (P, IF and N, IS). With both tendencies at work rather than one or the other, one would expect (as in the results actually obtained) the differences between the two success groups (P, IS and N, IS) to be increased while the differences between the two failure groups (N, IF and P, IF) would be decreased.

What basis is there for assuming a "positivity effect"? Other studies have indicated some such effect (e.g., 6). In addition, we conducted a small-scale experiment to see if it would occur. As part of another experiment (done with

[8] Significant interactions were also obtained between the note and the group experience of success or failure. The women rate the note-writers most favorably in the following order of experimental treatments: (1) P, GF; (2) P, GS; (3) N, GS; (4) N, GF. The interaction obtained with the male Ss is the reverse of that obtained with the women. They rate the note-writer most favorably in the following order of experimental treatments: (1) P, GS; (2) P, GF; (3) N, GF; (4) N, GS. Since the interactions between note and group treatment are not central to the theme of this presentation, they shall not be discussed further.

male and female college students) we had 12 Ss examine the positive notes and 12 Ss examine the negative notes used in the study reported here to form an impression of the note-writer; both types of notes were written in the same handwriting. They, then, rated the note-writer on the same rating scales as used in this experiment. The 24 Ss were not given any information about either the note-writer or the person to whom the note was addressed other than that they were both members of the same team and that the note-writer had been asked to write a note to the other person to indicate whether or not he wanted to be on the same team with the other person again. The results of this small-scale experiment are presented in Table 3. It is clear

TABLE 3

S's Evaluations of a Positive or Negative Note-Writer

	Positive Note-writer*	Negative Note-writer	"p" Value
1. "Evaluative" factor	2.4†	4.3	$<.05$
2. Like-dislike	4.0	6.0	$<.05$
3. Desirability as a friend	3.8	6.4	$<.05$
4. Desirability as a teammate	3.2	6.2	$<.05$
5. Team spirit	2.5	5.1	$<.05$
6. Intelligence	3.9	5.1	$<.05$
7. "Effectiveness" factor	3.0	3.5	N.S.
8. "Activity" factor	2.5	3.1	N.S.
9. "Status" factor	3.5	3.8	N.S.
10. "Strength" factor	2.8	2.1	N.S.
11. "Warmth" factor	3.3	4.6	$<.05$
12. "Self-control" factor	2.8	3.9	N.S.

* For scales 2–6 the most favorable rating is at "1" and the least favorable is at "9"; for scales 1 and 7–12 the most favorable rating is at "1" and the least favorable at "7."

† For scales 2–6, $N = 22$; for scalse 1 and 7–12, $N = 24$.

that there is a striking tendency for the Ss to evaluate a positive note more favorably than a negative note. These data validate the assumption of a "positivity effect" and, hence, give credibility to the interpretation of the results as supporting our original hypothesis.

Examination of the other data presented in Table 2 indicates that essentially the same results are obtained for almost all the other measures as were obtained for the more purely evaluative measures.[9] This suggests that the Ss responded to the note-writers in a relatively undifferentiated, evaluative manner; their evaluations primarily being determined by the "positivity

[9] For the men, statistically significant results were obtained only on the following measures: "evaluative" factor, rating of "like-dislike," and rating of "team spirit." The other measures showed the same trends but did not reach the .05 level of significance.

effect" and by "the tendency for cognitive balance" rather than by the specific characteristics that were rated. However, exceptions to the preceding statement should be noted. On the measures which presumably relate to task performance, i.e., the "effectiveness" factor, the "activity" factor, and the rating of intelligence, it is apparent that the S's own performance influenced his rating of the note-writer. Here, the more favorable evaluations by the Ss who experienced failure rather than success is explainable by assuming that the S's score relative to that of the note-writer indicated whether he was slower or faster, more effective or less effective, more intelligent or less intelligent, and that the Ss used this performance indicator in making his rating of the relevant characteristics.

TABLE 4

S's Evaluation of Herself after Receiving the Note

		Note	
		P*	N
Individual Performance	IS	2.1†‡	2.7
	IF	6.3	5.4

* See footnote *, Table 2.

† The scale has its most favorable rating at "1" and its least favorable rating at "7."

‡ Total N = 112. Both the Individual Performance and the Note/Individual Performance interaction were statistically significant ($p < .05$).

The Influence of the Note upon Self-evaluation

Although we had not deliberately planned to obtain a measure of the Ss' self-evaluation, after they had been exposed to the note, in fact we did so. The Ss filled out the self-ratings on the Semantic Differential after they received the note under instructions to rate how they felt when they learned their scores and the score of the other—i.e., under instructions which presumably would minimize the effects of the note. Nevertheless, for the women [10] there is a significant interaction effect between the note and individual success-failure on the scales that are most heavily weighted with the evaluative factor ("good-bad," "nice-awful"). The Ss who received a success treatment evaluated themselves more favorably if they were sent a positive rather than a negative note; for the women who received a failure treatment, the reverse is true. They evaluated themselves more favorably if they received a negative note. (See Table 4.) It is clear that it is not the positive note per se which raises self-evaluation nor the negative note per se which leads to self-depreca-

[10] The interaction effect is not quite significant at the .05 level for the men.

tion. Rather the effects of the evaluation made by the note-writer appear to be determined by their consistency or inconsistency with the S's self-evaluation. While it is impossible from our data to assert whether inconsistency produces a negative effect or whether consistency produces a positive effect or whether both effects occur, it is apparent that relative to one another inconsistent evaluations are more likely to have negative consequences than consistent evaluations.

The Effects of Initial Self-esteem and Actual Performance

Prior to the experiment we obtained a measure of the S's self-esteem and during the experiment we obtained measures of the Ss' actual performance on the tasks they were given. These measures were used to see whether or not the S's ratings of self or of the other were influenced by her self-esteem or by her actual performance. The correlations between the S's actual performance and her ratings of her performance did not differ from zero significantly. There were also no significant correlations between the S's actual performance and her evaluations of the note-writer. Likewise, an analysis of variance based upon a normalized form of the regression coefficients between the S's actual performance and her ratings of her performance showed no significant effects of the experimental treatments, or of their interactions.[11] These results suggest that the S's feelings of accomplishment or lack of accomplishment as she worked were not influential in determining her evaluations but rather that her evaluations of her performance were determined instead by the information supplied by the experimenter.

Similarly, the correlations between the measure of the S's initial self-esteem and her evaluation of her performance did not differ from zero significantly. Nor did an analysis of variance of a form of the regression coefficients between initial self-esteem and rating of one's own performance

[11] The formula for obtaining the scores to be employed in the analysis of variance based upon the regression coefficient was: $\sqrt{(\Sigma xy)^2/\Sigma x^2}$. The sign given to this expression was the same as that of the sum of cross-products. The error term used to test the significance of the main effects and interactions was obtained from the formula:

$$\frac{\Sigma y^2_w - \sum_1^g (\Sigma xy)^2/\Sigma x^2}{N-2g}$$

Σy^2_w refers to the within sum of squares for the dependent variables; g refers to the number of experimental conditions; N refers to the total number of subjects. Associated with this error term are $N-2g$ degrees of freedom.

We are indebted to Dr. A. P. Dempster of the Statistics Department for suggesting this procedure.

reveal any significant effects of the experimental treatments or of their interactions. These results suggest that the S's evaluation of her own performance was not influenced by her initial self-esteem, but rather that it was independently determined by the information supplied by the experimenter. It is, of course, possible that the measure of self-esteem that we used was not a good one and that this is the reason for the lack of relationship.[12]

SUMMARY AND CONCLUSIONS

The experiment reported here originated from the view that many of the theories currently being advanced in the field of social psychology have within them a hidden assumption that the individual has a positive evaluation of himself. We have advanced the notion that rather different predictions from those customarily made would be made, if the individual has a negative rather than positive view of himself, from such theories as Heider's theory of cognitive balance, Festinger's theory of dissonance, and Homans' interaction theory. Thus, these theories would predict that an individual will tend to regard favorably those whose evaluations of him are favorable and tend to regard unfavorably those whose evaluations are negative. Implicit in such a prediction is the hidden assumption that the individual evaluates himself positively. If the opposite assumption is made, namely that the individual evaluates himself negatively, one would predict that he would tend to view negatively another who esteems him and tend to esteem those who view him negatively.

To test the foregoing notions, we created an experimental situation in which the S: (a) was led to evaluate her performances on group-relevant tasks either favorably or unfavorably; (b) was led to believe that her group either won or lost a contest with another group because of, or in spite of, her performance; and (c) was exposed to a favorable or unfavorable evaluation of her perform-

[12] However, there is some evidence to indicate that the S's initial self-esteem (as measured by the TDMH prior to the experiment) did influence her reactions to the note she received. Significant findings were obtained (for the women but not for the men) when analyses of variance were performed on the regression coefficients between self-esteem and self-evaluation (on the "evaluative" factor of the Semantic Differential) and between self-esteem and the rating of the other on the "evaluative" factor. These findings may be summarized as follows:

(1) The higher the S's self-esteem the more favorably she rated herself (particularly if she had done well) after receiving a positive note; the higher her self-esteem the more unfavorably she rated herself (particularly if she had done poorly) after receiving a negative note. These results hold for the Ss who were on a winning rather than losing team.

(2) The higher the S's self-esteem the more favorably she rated the negative note-writer (particularly if the S had experienced individual and group failure).

These findings do not lend themselves to any simple, unequivocal interpretation.

ance and desirability as a teammate by a note which was presumably from another member of her group.

The S filled out questionnaires at appropriate points, which provided data concerning the S's evaluation of herself and her evaluation of the teammate who evaluated her. A $2\times2\times2$ factorial design was employed for female Ss, with 14 Ss in each cell; the design was replicated for male Ss with 6 Ss in each cell.

The results of the experiment may be summarized as follows:

1. Experimentally, we were able to induce the Ss to evaluate their own performances favorably or unfavorably by the information we gave them about their relative performances compared with others. The Ss' evaluations of their performances were uncorrelated with measures of their actual performance or with a measure of self-esteem obtained prior to the experiment. These results indicate that the Ss' evaluations of their performances were independently determined by the experimental manipulations.

2. The Ss' evaluation of their own performances strongly influenced their reactions to evaluations they received from another. The Ss who received negative notes evaluated the note-writers very unfavorably if they (i.e., the Ss) thought well of what they themselves had done, but rather favorably if they thought poorly of their own performances. The Ss who received positive notes evaluated the note-writers favorably, but more favorably if they thought well rather than poorly of their own performances.

3. Holding constant the Ss' evaluation of their own performances, there was a clear tendency for the Ss to respond more favorably to positive rather than negative evaluations from another. This we called "the positivity effect."

4. Holding constant the "positivity effect," there was a clear tendency for the Ss to respond more favorably to evaluations from another which were consistent, rather than inconsistent, with their own evaluations of their performances. These results support our initial hypothesis of a "cognitive balance effect."

5. The Ss' evaluations of themselves were more favorable if they received notes which were consistent, rather than inconsistent, with their own evaluations. Thus, the Ss who thought poorly of their performances evaluated themselves more favorably after receiving a negative note than a positive note. Similarly, the Ss who thought well of their performances evaluated themselves more favorably after receiving a positive note than a negative note.

Manuscript received: October 3, 1958
Revised manuscript received: February 16, 1959
Morton Deutsch
Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey

REFERENCES

1. Cooley, C. H., *Human Nature and the Social Order,* New York: Scribner's, 1902.
2. Deutsch, M., *Conditions Affecting Cooperation,* Final Technical Report for the Office of Naval Research, Contract NONR-285(10), February 1957.
3. Festinger, L., *Theory of Cognitive Dissonance,* Evanston, Ill.: Row, Peterson, 1957.
4. Heider, F., "Attitudes and Cognitive Organization," *Journal of Psychology,* 1946, 21, 107–112.
5. Heider, F., *The Psychology of Interpersonal Relations,* New York: Wiley, 1958.
6. Homans, G. C., *The Human Group,* New York: Harcourt Brace, 1950.
7. Jordan, N., "Behavioral Forces that Are a Function of Attitudes and Cognitive Organization," *Human Relations,* 1953, 6, 273–287.
8. Krech, D., and R. S. Crutchfield, *Theory and Problems of Social Pschology,* New York: McGraw-Hill, 1948.
9. Lecky, P., *Self-consistency: A Theory of Personality,* New York: Island Press, 1945.
10. Mead, G. H., *Mind, Self, and Society,* Chicago: University of Chicago Press, 1939.
11. Osgood, C. E., G. J. Suci, and P. H. Tannenbaum, *The Measurement of Meaning,* Urbana, Ill.: University of Illinois Press, 1957.
12. *Manual for the TDMH Self-Concept Scale* (mimeographed), Tennessee Department of Mental Health.
13. Selltiz, C., M. Jahoda, M. Deutsch, and S. W. Cook, *Research Methods in Social Relations,* rev. ed., New York: Holt, forthcoming.

Symptom Manifestations as a Function of Situational Press: A Demonstration in Socialization [1]

WILLIAM R. ROSENGREN,[2] *Emma Pendleton Bradley Hospital*

The maintenance of acting-out children in open residential treatment centers has proved to be a difficult task for psychiatric institutions (1, 7). By his very nature, the acting-out child appears ill-suited to an open residential setting. His impulsivity, his inability to engage in reciprocal relations, his constant desire for immediate gratifications, his frequently violent and spontaneous hostility, his short attention span—such attributes make him difficult to maintain in a residential setting. Yet, in its extreme form, the acting-out syndrome apparently cannot be tolerated outside of an institution. Therein lies the dilemma.

The Emma Pendleton Bradley Hospital in Riverside, Rhode Island, is currently engaged in a program of research to evaluate the effectiveness of various milieu therapies for acting-out children within an open residential treatment hospital for emotionally disturbed children. Part of this four-year study is concerned with the design of a series of special group activities for the acting-out child. These activities are to be of such nature that the structure of the activity situations themselves will tend to reduce the inappropriate manifestation of the acting-out syndrome. This paper reports an analysis of one such activity.

METHOD

Subjects

The subjects of this study were nine acting-out boys, ranging in age between eleven and thirteen, all of whom manifested the same basic symptomatology, and who had received clinical diagnoses of Passive-Aggressive Personalities—Aggressive Type. Typical complaints for which they were initially referred to the Bradley Hospital included such behaviors as inability to maintain lasting relationships with other persons, lack of affect, aggressiveness, impulsivity, lack of concentration, short attention span, and so forth.

[1] This study is part of a four-year investigation in social psychiatry which is supported by a grant from the National Institute of Mental Health, the National Institute of Health, United States Public Health Service.

[2] The author wishes to extend grateful appreciation to Carole Benson, Donald Cohen, M.S.W., and Norman Finkel, M.S.W., whose careful preparation of the process records, from which the data herein reported were abstracted, was an irreplaceable aid in this study.

Procedure

Beginning in mid-October of 1957 the nine boys were introduced to a series of planned cooking sessions in the hospital kitchen. Cooking may be regarded as a feminine activity in our culture and might, therefore, be considered an inappropriate therapy intervention for boys. It was selected, however, partially for that very reason. The patients involved tend to have overidentified with models of extreme masculinity—a fact which accounts, in part, for their hyperaggressiveness and their seeming difficulties in getting along with women. It was thought, therefore, that a series of experiences in a feminine activity would have less dysfunctional consequences for them than might be the case with undisturbed boys of their age. Cooking a meal, as a set of situations, was conceptualized as exerting a number of modest situational presses upon their symptomatology. In order to complete the task of preparing a reasonably satisfying meal, the boys had to engage in prior planning and preparation, they had to engage in considerable cooperative action in the preparation of the meal, they had to schedule various steps in the cooking processes, they had to be relatively inactive within limited physical boundaries, and so forth. These are all tasks which acting-out children have difficulty performing without spontaneous impulse break-through and violent verbal and physical aggression. Nonetheless, it was felt that the inherent satisfactions involved in the cooking activity itself would tend to offset the discomforts accruing from the presses upon their acting-out character which were imposed by the nature of the task-situations involved.

The adult leaders of the group maintained process record accounts of each cooking session. These reports were written systematically from mid-October of 1957 until late March of the following year. The sessions were then arbitrarily divided into a group of four "early sessions" which occurred between late October and early February, and a second group of four "late sessions" which took place between late February and late March.

The model consisted, therefore, of the meal cooking situations as independent variable, the overt behavior of the boys as dependent variable, with the passage of time as intervening variable. Time appears to be basic to any such patient changes—their behavior patterns being generally less labile or otherwise less susceptible to immediate situational press than would be true for more adjusted persons.

The descriptive accounts were then subjected to a content analysis employing a set of a priori categories. The content analysis was made in terms of five major dimensions of behavior:

1. Relevancy of behavior
2. Suggesting, supporting, and objecting behavior

3. Interpersonal interaction patterns
4. Applications of external controls by adult leader
5. Indices of group solidarity

For each individual within the group, frequency distributions were made for each major dimension and subcategories therein. A rating of "High" was given to a boy if he ranked one standard deviation or more above the mean on any behavioral category, and "Low" if he ranked one standard deviation or more below the mean. As an indication of significant change from the early sessions to the late sessions, Chi-Squared or Fisher's Exact Test or Probability was used, depending upon the frequencies involved.

PATTERNS IN EARLY SESSIONS

Relevancy of Behavior

A dichotomy distinguishing relevant behavior from irrelevant behavior was used in the analysis. Relevant behavior included all identifiable actions on the part of individual subjects which bore a direct relationship to the completion of the task of cooking. Irrelevant behavior included all actions which indicated concern with goals and satisfactions other than the completion of the task at hand. The category of relevant behavior includes, therefore, such gross actions as suggesting how best to prepare pizza pie, cutting up lettuce for salad, setting the table, and so forth. Irrelevant behavior includes such behaviors as a spontaneous game of tag, eating apples rather than slicing them for a pie, and so forth. It is to be noted that children of the acting-out type are prone toward considerable irrelevant behavior in most situations.

In the early sessions, just two individuals engaged in high relevancy, while three individuals engaged in high irrelevancy. The remaining four boys were about equal in regard to relevancy and irrelevancy. In terms of the contributions made by individuals, the early sessions were characterized as being pulled toward goal-directed activity by two individuals, and toward random and irrelevant behavior by three individuals. And of the total identifiable instances of relevancy and irrelevancy, 65 per cent were of a relevant nature while 35 per cent were of an irrelevant nature.

Suggesting, Supporting, and Objecting

Each of these major categories were subdivided into actions taken in a conciliatory fashion as opposed to those taken in an aggressive fashion. A suggestion, support of a suggestion, or an objection was understood to be aggressive in nature if the apparent consequence of the act was to depreciate the values and status of another person—to humiliate or degrade. Conversely, such actions were understood to be conciliatory in nature if there was no

apparent resulting humiliation, degradation, or depreciation of other persons. Typical acting-out behavior is aggressive rather than conciliatory.

In the early sessions, there was one high suggestor among the nine boys. This boy also happened to be the indigenous leader in the group. Supporting behavior was manifested primarily by one individual who happened to be the roommate of the high suggestor. On the other hand, four of the nine boys ranked high on objecting. Of the total behaviors classifiable under these rubrics, 26 per cent were suggestions, 32 per cent were supports of suggestions, and 46 per cent were objections. And of the total suggestions which were offered, 39 per cent were aggressive and 61 per cent were conciliatory. Of the total supportive actions, 11 per cent were aggressive and 89 per cent were conciliatory. And of the total objections which were made, 84 per cent were aggressive and 16 per cent were conciliatory. The tendency in the early sessions was for the suggesting and supporting roles to be enacted in a conciliatory fashion, and for the objecting role to be played in a highly aggressive fashion. However, the seeming pattern of conciliatory suggesting is somewhat spurious in view of the fact that just one person occupied the suggestor role, and this boy enacted it in a highly aggressive fashion.

Interaction Patterns

Separate codings were made for child-to-child, child-to-leader, and leader-to-child interactions. Each of these major interaction patterns were then subdivided into impulsive as compared with deliberative interactions. An interaction was understood to be impulsive in nature if it appeared to stem not from thoughtful reflection and consideration on the part of the actor, but, rather, from an inability to withstand the stresses of a situation involving the need to adjust to reciprocal expectations and normative behavior. Deliberative interactions were understood to be the reverse of impulsive actions. From the point of view of the clinical syndrome, it was expected that the boys would be low in child-to-child interactions, and generally quite high on all patterns for impulsive actions.

Of the total identifiable person-to-person interactions in the early sessions, 18 per cent were child-to-child, 46 per cent were child-to-leader, and 36 per cent were leader-to-child. In the early sessions there was a significant lack of child-to-child interactions ($\chi^2=.001$). In terms of total interactions by individuals within the group, the interpersonal field was dominated by one individual, who also was the indigenous leader as well as the high suggestor. Two individuals were significantly low in all types of interactions. There was a tendency for individuals with high interaction rates to have a correspondingly high ratio of deliberative actions to impulsive actions ($\chi^2=.05$). And con-

versely, the boys with low total interaction patterns tended to have a high ratio of impulsive actions to deliberative actions (χ^2=.01).

In the early sessions, there was a significantly high percentage of impulsive child-to-child interactions as compared to deliberative child-to-child interaction (χ^2=.05). And similarly, there was significantly more impulsive child-to-leader interactions than deliberative (χ^2=.01). Leader-to-child interactions, however, tended to be significantly deliberative in nature (χ^2=.001). Moreover, where a child interacted deliberatively with another child, the response tended to be deliberative. And where a child interacted impulsively with another child, the response tended to be impulsive. But where the adult leader interacted deliberatively with a boy, the response tended to be impulsive rather than deliberative—a pattern of lack of reciprocity that is to be expected of acting-out children.

Application of External Controls by Adult

The usual method of controlling children who exceed the limits of acceptable behavior in the hospital is to remove them from the situation in which they are involved and have them sit quietly by themselves until they can regain the use of available internal controls. During the early sessions control through "sitting" was imposed on an average of twice per session. Three separate classifications were made for reasons for the application of the control:

1. Disruptive behavior on the part of a boy relative to another boy.
2. Disruptive behavior on the part of a boy relative to the adult leader.
3. Spontaneous impulse break-through on the part of a boy with no identifiable interactional setting.

In the early sessions, 65 per cent of the controls were applied for spontaneous impulse break-through, and the remainder for disputes with the leader. There were no instances of controls for disputes between the boys themselves. And this was in spite of the fact that the ratio between deliberative and impulsive behavior was about the same for child-to-child interactions as it was for child-to-leader interactions.

Group Solidarity

One relatively simple criterion was used to indicate either the presence or absence of group solidarity—praise or criticism of others during the cooking session. Because of the generally unsocialized character of the acting-out child, and because previous observations of the boys seemed to reveal that much of their activity was nothing more than parallel play, it was felt that there would be only slight indications of group solidarity, as phrased in these

terms. Of the total indications of solidarity, 28 per cent indicated a presence of solidarity while 72 per cent indicated a positive lack of solidarity. There was, in other words, considerably more criticism of others than there was praise. The inference might be drawn from these data that the pleasures and satisfactions which the boys derived from the early sessions were of a distinctly autonomous nature, rather than accruing from an orientation toward the enjoyment of a reciprocal task.

Summary of the Early Sessions

In general, the early cooking sessions were characterized by the following patterns:

1. High frequency of irrelevant behavior as compared with relevant behavior.
2. High relevancy manifested by two individuals in a group of nine.
3. High objecting behavior relative to suggesting and supporting behavior.
4. High aggressive suggesting and low conciliatory suggesting.
5. High aggressive objecting and low conciliatory objecting.
6. Low child-to-child interactions.
7. Dominance of the interactional field by one individual—an individual who also wielded general indigenous power.
8. Positive association between high interaction rates and high deliberative behavior.
9. Positive association between low interaction rates and high impulsive behavior.
10. Interactions from child-to-child and from child-to-leader tended to be impulsive rather than deliberative.
11. Application of external controls primarily for spontaneous impulse break-through.
12. General indications of a lack of group solidarity.

Most of these patterns are what might be expected of patients of the acting-out type. If the sequential experience was, indeed, exerting a meaningful situational press upon the acting-out symptomatology, an analysis of the late sessions should reveal a reversal of at least some of the trends which were found to be present in the early sessions.

PATTERNS IN THE LATE SESSIONS

Relevancy of Behavior

There was some absolute change in the predicted direction, but it was not a significant change. In the early session 65 per cent of the goal-directed activity was relevant while in the late sessions 70 per cent of such behaviors

were relevant. This does not indicate a statistically significant change. In terms of individuals, however, a meaningful change did occur between the two groups of meetings. In the early sessions highly relevant behavior was manifest by two individuals, but in the late sessions three persons behaved in a highly relevant fashion. In the early sessions three boys behaved predominantly irrelevantly, but in the late sessions just one boy did. Although there was not an over-all tendency toward greater relevancy in the late sessions, there were more boys exerting pressures toward goal-directed activity and fewer individuals pressuring toward irrelevancy.

Suggesting, Supporting, and Objecting

As contrasted with the early sessions, no one individual dominated the suggesting role in the late sessions. In fact, the boys almost disappeared in this category, while the adult leaders assumed a much more important place in this respect. There was, however, a tendency in the late sessions toward more conciliatory suggesting and less aggressive suggesting on the part of the boys ($\chi^2=.05$), a pattern that was just the reverse in the early sessions. The most significant change occurred with respect to the manner in which objections were made. In the early meetings there was significantly more aggressive objecting than conciliatory objecting. But in the late sessions, conciliatory objecting far outweighed aggressive objecting ($\chi^2=.001$).

Interaction Patterns

Of the total identifiable person-to-person interaction patterns in the late meetings, child-to-child accounted for 11 per cent, child-to-leader for 41 per cent, and leader-to-child for 48 per cent. These do not indicate significant changes. Low interaction rates between the boys were maintained in both the early and the late sessions. In the early meetings, however, there was significantly more impulsive than deliberative interaction between the boys. This pattern was reversed in the late meetings; deliberative interaction between the boys was significantly higher than impulsive interaction ($\chi^2=.01$). And the same basic change occurred with respect to interactions from the boys to the adult leaders. In the early meetings, child-to-leader interactions were typically impulsive. In the late sessions, however, there were definitely more deliberative than impulsive interactions toward the leader ($\chi^2=.01$). And this is in spite of the fact that the adult leaders had been working with the boys for nearly a year and had no recollection of such changes having previously occurred in any other situation.

With respect to leader-to-child interactions, the pattern remained the same in both the early and the late sessions—significantly more deliberative than

impulsive actions. In the late sessions, however, the association between the extent of interaction and the kind of interaction disappeared. In the early cooking meetings, those boys who had high interaction rates tended to behave in a characteristically deliberative fashion, while those with low interaction rates tended to behave in a characteristically impulsive fashion. These patterns did not exist in the late meetings.

Application of External Controls by Adult

There was a significant change in the indices of group solidarity between the two groups of sessions. In the early meetings, 72 per cent of the pertinent behaviors indicated an absence of solidarity. But in the late sessions, the figure was reduced to about 15 per cent. Indices of the presence of solidarity increased from 28 per cent in the early meetings to about 85 per cent in the late sessions. This was a significant change ($\chi^2=.01$).

CONCLUSIONS

It was found that two patterns which are typically acting-out in nature were maintained in both the early and the late sessions. First, the boys tended to engage in a great deal of irrelevant behavior. And second, there was a significantly low rate of child-to-child interaction, of a meaningful nature, in both periods. However, a number of typically acting-out behaviors which were found to characterize the early sessions were not present in the late sessions:

1. There were more individuals in the late sessions than in the early sessions pulling the group toward goal-directed activity and fewer individuals pulling the group toward random and irrelevant behavior.
2. The pattern of a high ratio of objecting behavior to suggesting and supporting behavior in the early sessions changed, in the late sessions, to a high ratio of suggesting and supporting behavior to objecting behavior.
3. The early sessions were characterized by high aggressive suggesting, while the late sessions by high conciliatory suggesting.
4. The early sessions were characterized by high aggressive objecting, but the late sessions by high conciliatory objecting.
5. Social interaction was dominated by just one of the nine boys in the early sessions, but in the late sessions significantly high interaction rates were shared by three of the nine boys.
6. In the early sessions, high interaction rates were associated with deliberative behavior and low interaction rates with impulsive behavior. In the late sessions, however, the association between extent of interaction and kind of interaction disappeared.

7. In the early sessions disciplinary me sures were applied primarily for spontaneous impulse break-through, but in he late sessions, instances of impulse break-through were significantly fewer.

8. In the early sessions, there were positive indications of a lack of solidarity, suggesting that the boys did not perceive the experience at a set of reciprocal relations; but in the late sessions there were significantly more indications of solidarity.

9. In the early sessions child-to-child and child-to-leader interactions tended to be highly impulsive. In the late sessions, however, these interactions tended to be significantly more deliberative.

10. In the early sessions, deliberative behavior on the part of the leaders tended to call forth impulsive behavior on the part of the boys. But in the late sessions interactions tended to be more reciprocal—deliberative behavior on the part of the leaders more frequently called forth deliberative behavior on the part of the boys.

11. In the early meetings reciprocity between particular boys appeared to be related to the general clique arrangements in the group. In the late sessions, however, reciprocities no longer appeared to be contingent upon ascribed status in the group.

In general it can be said that in the late sessions there were considerably fewer indications that the boys were translating their acting-out characteristics into overt action than was the case in the early sessions. Changes in the behavior patterns from the early meetings to the late meetings indicate that the boys were more responsive to social expectations as well as less psychologically acting-out, to the extent that such a distinction may legitimately be made.

DISCUSSION

It would seem that these data relate to the issue of heterogeneous versus homogeneous groupings in psychiatric institutions. Each residential policy rests upon a different set of postulates concerning the processes of milieu therapy.

The policy of heterogeneous grouping (grouping several individuals together with a variety of different emotional and mental disturbances) is apparently founded upon the principle of what might be called the "balanced group." From this point of view, the disturbed individual is presumably sensitized to his own inadequacies by living in a situation in which he perceives in other persons the characteristics which he lacks. Even patients who are heterogeneously grouped, however, do things together in common situational settings. They are all exposed, at one time or another, to the same role prescriptions accruing from the structure of a common social field. It would

seem, therefore, that some of the activities of the heterogeneous group would provide relearning experiences for some patients, but merely help to stabilize what may be learned disorders for others. What may be actually happening is that through heterogeneous grouping pathological desires are met as frequently as are therapeutic needs. The presence of an overly aggressive person in the same living group with an overly passive individual provides opportunities for the aggressive individual to be pathologically aggressive and for the passive individual to be pathologically passive, depending upon the patterning of the expectations to which they are exposed. It is difficult, in other words, to structure common role expectation situations which are therapeutic for both the overly aggressive and the overly passive at the same time.

Homogeneous grouping seems to rest, on the other hand, on the principle that behavior is a function of a transaction between the inner state of persons and the outer conditions in which they find themselves. That is, social processes consist not only of interactions between individuals but of interactions between individuals and the social field in which they are located.

It is in the context of homogeneous grouping that the concept of "situational press" is most meaningful. The concept of situational press may be understood to mean the structuring of situations in which the attainment of goals is facilitated by the adaptation to specified behavior patterns on the part of the persons involved. Thus, what might be a therapeutic situational press for one disorder might serve only to firm up existing symptom patterns for another.

Numerous studies of psychiatric institutions have demonstrated that the behavior of both patients and staff are a function of relationships between internal conditions and outer situations (10, 2, 8, 3, 4, 9, 5). This being true, both increases and decreases in overt manifestation of symptoms are potentially manipulable for purposes of both control and therapy. For the acting-out child, the problems of control are undoubtedly as profound as are the problems of therapy. One can hardly proceed or succeed without the other. Indeed, one might say that the acting-out child's need for therapy is directly related to his control problem. And the need for external controls can be either increased or decreased depending upon the repeated situational presses to which the child is exposed. Viewed historically, the emergence of the acting-out syndrome may be traced to a variety of early childhood experiences, with or without organic components (6). It seems, however, that over time the acting-out behavior patterns attain some measure of functional autonomy; they become habitual and adaptive in their own right. Through repeated experiences, the acting-out child seems to have learned that he can facilitate the attainment of desired goals through acting-out means. The

disturbance is both adaptive and habitual in that sense. And he has also learned to value only those goals which can be attained by acting-out means. It is adaptive and habitual in that sense as well. Viewed historically, the acting-out syndrome is a set of expressive defenses to past stresses; viewed a-historically, it is a set of instrumental means for the attainment of contemporary ends.

One of the chief tasks of milieu therapy for acting-out children is, then, to provide situations in which the acting-out child may learn to value a richer set of goals and, at the same time, learn that the attainment of goals may be more easily facilitated through nonacting-out means. To the extent that this may be done, it serves as an adjunct to individual psychotherapy and reduces the need for external controls. The latter, in fact, frequently appears only to reaffirm the acting-out child's perception of the world as punitive and hostile. It may be through thoughtful application of the concept of situational press that these kinds of relearnings may take place.

Manuscript received: August 27, 1958
Revised manuscript received: January 27, 1959
William R. Rosengren
Emma Pendleton Bradley Hospital
Riverside 15, Rhode Island

REFERENCES

1. Bettelheim, B., *Love Is Not Enough,* Glencoe, Ill.: The Free Press, 1950.
2. Caudill, W., *The Mental Hospital as a Small Society,* published for the Commonwealth Fund by Harvard University Press, Cambridge, Mass., 1958.
3. Caudill, W., and E. Stainbrook, "Some Covert Effects of Communication Difficulties in a Psychiatric Hospital," *Psychiatry,* 1954, 17, 27–40.
4. Devereaux, G., "The Social Structure of a Schizophrenic Ward and Its Therapeutic Fitness," *Journal of Clinical Psychopathology,* 1944, 6, 231–265.
5. Klemes, M. A., "The Therapeutic Effect of Group Morale in a Psychiatric Hospital Ward," *Bulletin of the Menninger Clinic,* 1951, 15, 58–63.
6. Laufer, M. W., E. Denhoff, and G. Solomons, "Hyperkinetic Impulse Disorder in Children's Behavior Problems," *Psychosomatic Medicine,* 1957, 19, 38–49.
7. Redl, F., and D. Wineman, *Children Who Hate,* Glencoe, Ill.: The Free Press, 1951.
8. Scher, J. M., "The Structured Ward: Research Method and Hypothesis in a Total Treatment Setting for Schizophrenia," *American Journal of Orthopsychiatry,* 1958, 28, 291–299.
9. Sheimo, S. L., J. Paynter, and S. A. Szurek, "Problems of Staff Interactions and Spontaneous Group Formations on a Children's Psychiatric Ward," *American Journal of Orthopsychiatry,* 1949, 19, 599–612.
10. Stanton, A. H., and M. S. Schwartz, *The Mental Hospital,* New York: Basic Books, Inc., 1954.

The Maintenance of a Learned Response in Controlled Interpersonal Conditions [1]

SEYMOUR ROSENBERG,[2] *University of Kansas*

There is a growing body of literature in which clinical and interpersonal phenomena are analyzed in terms of learning principles (2, 3, 8, 10, 13, 16). These analyses have been supplemented with experiments whose express purpose is the extension of reinforcement principles to complex psychological phenomena (9). Most of these experiments have focused upon studies of the reinforcing properties of interpersonal stimuli and upon the ability to control complex behavior, particularly verbal behavior, rather than upon interpersonal interaction, per se.

Social interaction, in these experiments, has typically been limited to linkages between a single subject and experimenter, or subject and stooge. The essential point is that the behavior of the experimenter or stooge is prescribed in advance according to certain a priori rules. Generalization from these experiments to two-person interactions is probably limited, therefore, to interpersonal situations where the response contingencies of one person of a dyad conform to these rules. Analysis of social interaction also requires studies where the behavior of neither person is fixed in advance. Such an extension would specify reinforcement rules for each of *two* (or more) Ss based upon the responses of these Ss. Although the use of a group product measure usually accompanies such extensions (e.g., 6, 7, 12), analysis of individual behavior as a dependent variable is not precluded by this extension of the independent variable.

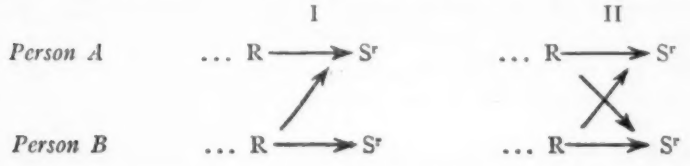
This study concerns the maintenance of a trial-and-error response or operant in a dyadic situation. A simple motor response was first learned by each individual in routine operant learning conditions. A range of response magnitudes was arbitrarily selected by the experimenter and designated the correct response. Reinforcement to a S during preliminary training was contingent only upon his performance of the response. The reinforcing event was a quantitative feedback delivered after each response with the "good" range clearly displayed to the S.

[1] This study was performed by the author at Lackland Air Force Base, Texas. This research was done under ARDC Project No. 7737, Task No. 27006, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] The author wishes to thank Dr. Robert L. Hall for his helpful suggestions in the conduct of this study.

This routine training was conducted separately but simultaneously for two individuals at a time. This procedure permitted the experimenter to change, without pause, to a social basis for delivering reinforcement to each person after preliminary individual training. Depending upon the experimental condition, the reinforcement values to one person were made to depend, in part or entirely, upon the response magnitudes of the other person. That is, for a person to continue to receive feedback in the "good" range, both his own and another person's responses (or only the other person's behavior) had to conform to a specified range of magnitudes.

Since reinforcement values in the dyadic conditions continued to be displayed to each person separately, it was possible to assign each of the individuals of a pair to a different degree of social contingency. Many such combinations are possible and the following diagrams illustrate two of them:



The value of person A's reinforcing stimulus, S^r, in both examples is contingent upon both his behavior and that of person B. In Example I, however, person B's S^r is independent of person A's behavior and is, therefore, asocial for him. Example I is termed a nonreciprocal linkage. Example II is a linkage in which the reinforcing contingencies are reciprocal, i.e., where the S^r's of both persons A and B are contingent on the other's behavior. Person B's S^r in Example II is contingent entirely upon person A's behavior. Recent studies by Hall (6), Rosenberg and Hall (12), and Sidowski *et al.* (14, 15) exemplify reciprocal linkages. The specification of reinforcement relationships must also include a denotation of the precise mathematical function or rules by which the operants will be combined and transformed into an individual's S^r. When these specifications are made for both individuals, we have constructed an interpersonal reward system and can study response changes of each individual as a function of this system.

PROCEDURE

Apparatus

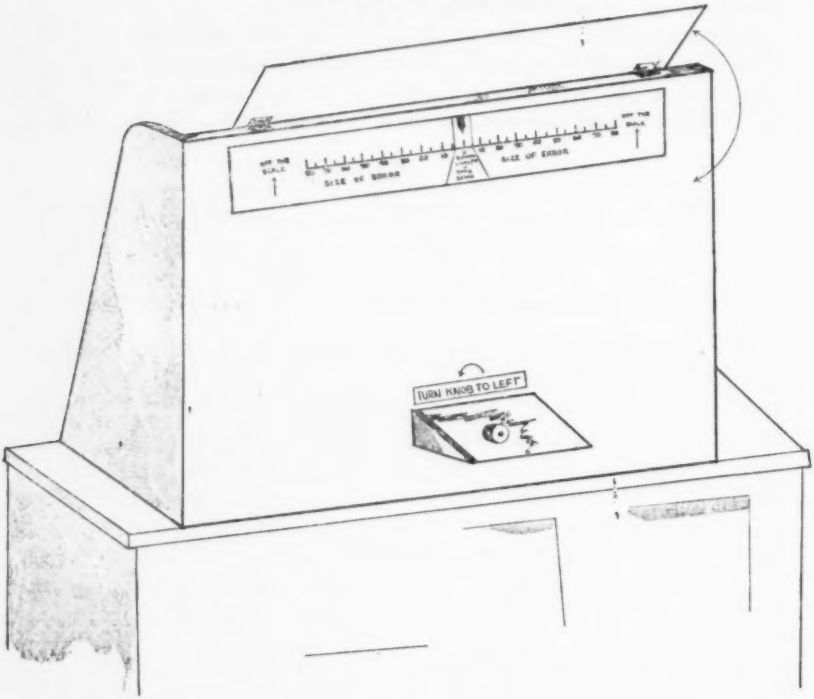
Two Ss [3] were seated in identical adjacent booths. The walls of the booth were sufficiently high so that Ss could not see each other. One wall of each

[3] All persons are termed Ss in the Procedure section. For data analysis it was necessary to designate one member of the pair "subject," and the other member "partner." Such

booth consisted of the panel and table as shown in Figure 1. Each S faced separate panels. A micrometer was mounted behind the panel so that the knob at the beginning of each trial projected ¼ inch from the face of the panel. The S could touch only the knob portion of the micrometer. He could turn the knob by reaching under an opaque shield (shown in Figure 1 as

FIGURE 1

A Sketch of the Panel Facing Each S.



partially cut away), but had no visual cues nor special tactual cues from the knob to indicate the amount it was turned. The knob could only be turned to the left. The E, seated on the other side of the panel, could read from the micrometer scale the amount the knob was turned. Also, on the panel in front of each S, a linear scale extending from zero in the center to 80 in each direction, indicated "size of error" to S. A movable pointer above the

designations were not made until all data were collected and, hence, the distinction is reserved for the Results and Discussion section.

scale displayed a score to the S after each trial. The scale was concealed by a hinged opaque cover during the Ss' response period and during a short delay before reinforcement. The scale also contained "off the scale" marks at each end. At the center of the scale an interval of 10 units was marked off by two red lines and was labeled "try to get pointer within red lines for good score." No cues were present on the scale to indicate whether error scores to the left of zero on the scale indicated too large or too small a response magnitude. This was counterbalanced by displaying excessive response magnitudes to the left of zero for half of the Ss and to the right of zero for the other half of the Ss.

Subjects

The Ss were male Air Force basic trainees whose modal age was about seventeen years. Trainees were assigned as a unit for use as Ss in psychological research but were excluded from the experiment if their Armed Forces Qualification Test (AFQT) score was below the 65th percentile rank, which is roughly equivalent to a Stanford Binet I.Q. of 108 (4). Ss were used in pairs, randomly assembled.

Seventy-nine Ss were given preliminary individual training on the response but only 54 Ss satisfied a predetermined criterion of response proficiency. This criterion will be described in the Preliminary Individual Training section below after the task has been described more fully. The Ss (25 in number) who failed to satisfy the criterion of learning during preliminary training were discarded as slow learners. It was assumed that, with sufficient trials, the 25 discarded Ss would have met the criterion and could have been used in the experiment proper. These Ss were simply discarded to avoid a lengthy preliminary training session. The 54 Ss who were retained for the experiment proper were given additional trials in one of the social interaction conditions.

Preliminary Individual Training

Each S was given the same instructions followed by 25 trials. During these 25 trials, the reinforcement values given to each S depended only upon the S's own response. The following are details of instructions, time during trials, and scoring.

After the two Ss were seated in adjacent booths, they were given tape-recorded instructions describing their task in the experiment. Ss were told that their job consisted of learning to turn the knob in a series of trials so as to minimize the size of the error displayed after each trial. The Ss were given no cues that they might be working as a "team." The wording of the instructions may be found in a previous study (12).

The procedure for the complete trial consisted of three automatically timed intervals: (a) a response period of 8 seconds, indicated to the Ss by a mild buzzer, during which the knob could be turned; (b) a delay of 10 seconds during which E recorded response magnitude and moved the scale pointer (still concealed from each S) to its appropriate value for that trial; (c) the presentation of knowledge of results by lifting the hinged shield and exposing the location of the pointer to the S. During this interval the micrometer was reset by E to its original position. The shield was lowered after 3 seconds and the response interval for a new trial was started.

Response magnitude for each S during each trial was read by the E directly from the micrometer scale. Twenty-five scale units equals one complete knob rotation. During preliminary training, a perfect score (goal score) was 100 micrometer units (four knob turns).

The criterion for retaining an S for the experiment proper was obtained from trials 21–25 as follows: The absolute error, i.e., deviation from 100 in either direction, obtained by an S on each of these five trials was summed. If this sum exceeded 100 micrometer units, the S was discarded. For example, if an S turned the knob the following amounts on the last five trials, 90, 110, 113, 95, 78, the corresponding absolute errors would be 10, 10, 13, 5, 22, the sum of which is 60; hence he would be retained. If an S turned the knob the following amounts, 90, 130, 63, 140, 100, his summed *absolute* error is 117 and he would be discarded from the experiment proper.

Experimental Trials

Each S who satisfied the criterion of individual proficiency just described was assigned to one of three reinforcement conditions and given 50 additional trials in this condition. If the criterion was met by each of the two Ss who were run simultaneously in preliminary training, they were assigned to conditions which required the simultaneous presence of both persons. These are the reciprocal linkages. If only one S satisfied the criterion, he was assigned to a nonreciprocal linkage and paired with a person whose data were collected at another time. Nonreciprocal linkages do not require the simultaneous presence of both persons so long as the data from one member of the pair are available when the second member is run. Precautions were always taken to avoid giving any cues to any S who was retained that he had changed from preliminary individual training to the experimental trials.

The three reinforcement conditions to which an S could be assigned for his experimental trials are: (a) An S^r value on each trial based upon the magnitude of his response only, termed "direct" values. This condition is exactly the same as that used for all Ss during preliminary individual training. A perfect score is obtainable by S if he turns his knob 100 micrometer units. (b) An S^r value on each trial which combines in a simple additive

function the magnitude of his own response with that of a second S, termed "confounded" values. In this condition a score to the S was computed by averaging the response magnitudes of the two persons. An *average* of 100 micrometer units was a perfect score. (c) An S^r value on each trial based upon the magnitude of another's response only, termed "other's" values. An S receives a perfect score when another S makes four knob turns or 100 micrometer units. Note that when both Ss are first trained to approximate closely a standard response of 100 units, the shift to "confounded" or "other's" values may not be discernible to S.

These three conditions comprise the two extremes and middle point of a continuum. This continuum specifies the fractional weight given to a S's response relative to the fractional weight given to his partner's response in the S^r to the S. When the S's response weight is 1.00 and his partner's response weight is .00, the S^r value to the S is "direct" (one extreme of continuum). When S's response weight is .00 and the partner's is 1.00, the resultant S^r value is "other's" (the other extreme of the continuum) for the S. An equal weighting of .50 for each person's response yields a "confounded" S^r value (the middle point of the continuum).

Since the S and his partner may receive different S^r's, there are nine possible combinations of these continuum values, i.e., one of three points to the S and one of three points to the partner. Three dyads were assigned to each of the nine combinations.[4]

RESULTS AND DISCUSSION

For purposes of analysis, a distinction was made between the experimental condition of a *Subject,* S, and that of his *Partner,* P. One member of each of the 27 pairs of persons was randomly selected as the S in the analyses and the second member became the *Partner,* P. Only dependent variables associated with the S during the 50 experimental trials are presented below.[5]

[4] From the standpoint of the Ss, the three combinations in which S is on "direct" values are the same. The three associated conditions of the partner can have no differential effects on the S's responses as long as the latter is receiving direct S^r's. This experimental design was chosen, however, to satisfy an analysis of variance which would clearly and conveniently distinguish effects on S's response which are attributable to his own condition from effects which are attributable to his partner's condition. Designs which collapsed the three combinations in which S is on "direct" values would present awkward analysis problems.

[5] Ps' response data were also analyzed and the results compared informally with analyses of Ss' response data. The "replication" was judged to be very good but is not presented as a formal replication because the two sets of data were not considered completely independent of each other. That is, there exists an experimental dependency between the two members of a dyad in many of the combinations of experimental conditions. The "replication" analyses are summarized in the Appendix for readers who wish to compare the two sets of data.

Two dependent variables were computed for each of the 27 persons designated as Ss. These measures are: (a) "response accuracy," $RA = \log (|x| + 1)$, where x is the deviation of response magnitude from 100 in micrometer units (goal score in preliminary training) of S on a trial; (b) "stimulus magnitude," $S^rM = \log (|E| + 1)$, where E is the numerical value of the error displayed to an S on a trial. The second measure is also viewed as a response measure. It is a measure of response "function," i.e., a definition of a response in terms of the procurement of reinforcement. The logarithmic transformations reduced skewness. The value, one, was added to the absolute value in each measure to avoid the log of zero, which is not a finite number.

Each measure was subjected to a separate analysis of variance. The design, the same for both measures, permitted analysis of the effects of reinforcement conditions of S and the effects of reinforcement conditions of P on Ss' two measures. Thus, there were nine cells arranged in a 3×3 matrix. Rows are Ss' conditions and columns are Ps' conditions. Each cell in this matrix contained data of three Ss. The statistical design and the results are illustrated in Figure 2. The nine graphs are arranged in rows and columns which correspond with S's and P's conditions, respectively, in the analysis of variance design. Each graph includes a plot of the RA (solid line) and the concomitant S^rM (dashed line) during the last block of preliminary trials and the ten experimental trial blocks. The statistical analysis, however, retained the integrity of each of the 50 experimental trials and did not include data from the preliminary trials.[6]

Analyses of variance are summarized in Table 1. The F ratios are based on error terms regarded as most appropriate for generalizing the effects over individuals. Trend effects over trials were studied by extracting orthogonal polynomials and were limited to linear and quadratic components (5). Extraction of these two orthogonal components supplement the test for over-all trials effect since the latter is too general a test of trend and fails to distinguish simple trend types.

In general, the RA analysis indicates that only reinforcement conditions of S, himself, result in significant differences in this measure. These differences occur both in average effects over trials (Source 1), in trend effects (Sources

[6] Plotting blocks of five trials, while presenting a clear picture of trend differences, also dilutes dramatic changes in simple trends which may take place within a block of five trials. This problem occurs particularly during the first experimental trial block. Data from the last preliminary trial block were used, therefore, to illustrate large initial changes. In a sense, trial one in the experimental period would show the same effects as the last trials of the pretraining period as long as little learning occurred in the last five pretraining trials. This seemed to be the case when these trials (21–25) were inspected qualitatively. In this sense, therefore, the graphs depict well the data analyzed statistically.

TABLE 1

Summary of Statistical Analyses of the Individual Response and Reinforcement Stimulus Measures

	df of Each Analysis		Response Accuracy RA = log (|*x*| + 1)		Stimulus Magnitude S'M = log (|*E*| + 1)		Error Term for *F*
			Mean Sq.	*F*	Mean Sq.	*F*	
1. Rows (Ss' conds.)	2		582,170	15.98‡	363,169	20.72‡	4
2. Columns (Ps' conds.)	2		37,310	1.02	134,382	7.67†	4
3. Rows × columns	4		72,192	1.98	141,624	8.08‡	4
4. Within cells	18		36,433		17,529		
5. Trials	49		3,365		1,975		
6. Rows × trials	98		2,038	1.55†	1,798	1.33*	9
a. Linear		2	41,867	6.64†	26,600	3.11	9a
b. Quadratic		2	9,688	4.98*	8,312	2.60	9b
c. Residual		94	1,028	.86	16,010	13.78‡	9c
7. Columns × trials	98		1,410	1.07	1,607	1.19	9
a. Linear		2	7,674	1.22	13,182	1.54	9a
b. Quadratic		2	6,954	3.58*	7,342	2.30	9b
c. Residual		94	1,158	.97	1,239	1.07	9c
8. Rows × columns × trials	196		1,498	1.14	1,136	.84	9
a. Linear		4	18,058	2.86*	8,425	.98	9a
b. Quadratic		4	1,202	.62	1,628	.51	9b
c. Residual		188	1,152	.96	971	.84	9c
9. Within cells × trials	882		1,316		1,352		
a. Linear		18	6,309		8,566		
b. Quadratic		18	1,944		3,198		
c. Residual		846	1,196		1,162		

* = $P < .05$.
† = $P < .01$.
‡ = $P < .001$.

6, 6a, 6b), and are significant at the .054 level in the second order interaction (Source 8a).

Figure 2, viewed in terms of row effects, illustrates very clearly the direction of the differences in RA as a function of Ss' condition differences. The solid lines, RA, show virtually no deterioration, nor much improvement, when the preliminary training condition is retained during the experimental trials (top row of graphs). This condition is, in a sense, a "base line" or control condition. By contrast the social conditions result in rapid and extensive deterioration. The results for row effects can be rather simply summarized as indicating S's response accuracy is maintained to the extent that his reinforcement values are contingent upon his own behavior.

Column effects on RA are suggested by Figure 2, i.e., S's behavior deteriorates to the extent that it is contained in P's reinforcement values. The effect is considerably reduced in the analysis, however, and is most noticeable on the graph when S is in the confounded condition. The "replication" using P's response data *does* show a significant column effect. A reduced column effect would be expected, to some extent, because it is known that P's condition will make no difference in the top row (when S is on direct). This would, in turn, suggest an increased row $\times$ column interaction (Source 3) in Table 1 which also failed to reach significance. The value of extensive discussion of this apparent tendency is dubious in view of the lack of significant interaction and of the inconsistency of significance levels between nonindependent sets of response data.

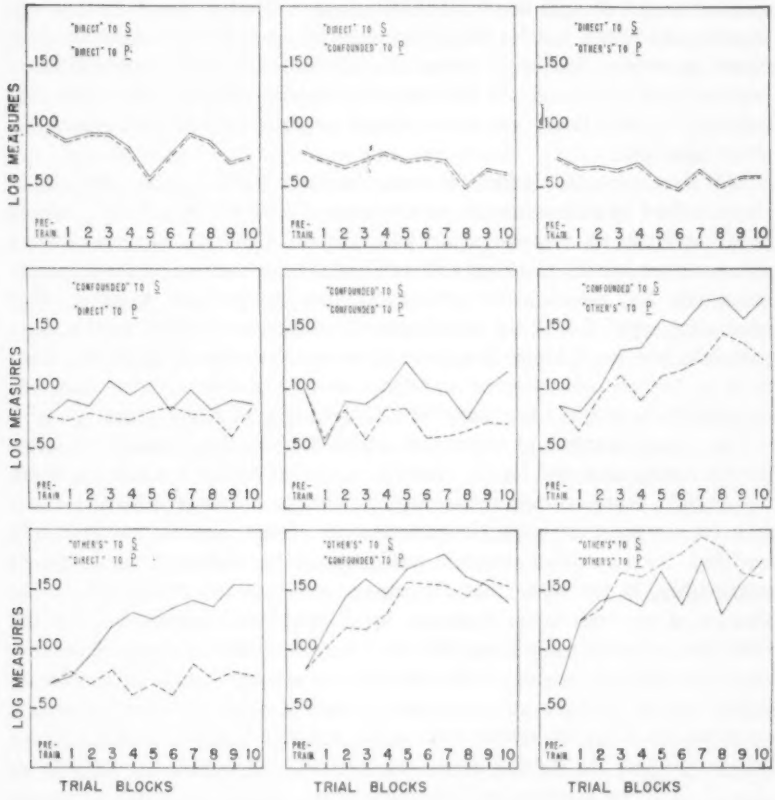
The nature of the second-order interaction effects on RA is illustrated in Figure 2 since the trends in each of the nine cells have been graphed separately. These second-order interaction effects do not appear to be complex or irregular. They may also be summarized in terms of the proportion that each person contributes to his reinforcement values. Figure 2 suggests that the second-order interaction effects of S's and P's reinforcement conditions can be described as an exaggeration in linear rate (over and above main effects) of deterioration of RA. This takes place as both individuals' reinforcement values become more contingent upon each other's behavior and less upon their own behavior. The precise functional relationship to express the interaction was not determined.

The analysis of S^rM or "response function" in Table 1 does not show the same significant effects as those obtained with RA. Effects averaged over trials are all significant (Sources 1, 2, 3). There is also some suggestion in the analyses that S^rM *trends* vary as a function of the reinforcement condition to an S (Sources in 6) but there is no consistent picture. Indeed, the highly significant residual trend (6c) is difficult to interpret at all. However, this fact along with the large but not quite significant linear and quadratic trends

(Sources 6a, 6b) may suggest that the most appropriate curve comparison would be one involving all orthogonal components. Figure 2 summarizes the results of S^rM in the nine cells with the dashed line.

FIGURE 2

The solid line in each graph is mean "response accuracy" (RA) and the dashed line is mean "stimulus magnitude" (S^rM) associated with the Ss in each of the nine reinforcement conditions. Values are shown for the last preliminary trial block and for the experimental trial blocks.



The S^rM curves seem to be quite different from those found for RA in some cells as both the graphs and statistical analyses testify. Thus, the amount of covariation between these two aspects of the behavior would not seem to be adequate to provide either a simple summary of both dependent

variables or to permit an "explanation" of one variable in terms of the other. In general, three kinds of trend relationships exist between RA and S^rM. With more intermediate points in row and columns these three kinds of trend relationships would probably "fuse" into each other.

One relationship is represented by the condition in which an S receives direct values (top row). RA and S^rM curves are, of course, the same and generally appear to be horizontal straight lines or, perhaps, lines with a slight negative slope.

A second general relationship exists where the S^rM values generally remain as a horizontal straight line but RA deteriorates. That is, the S continues to receive good scores but his response magnitude is gradually changing. This occurs in varying degrees in three cells, *confounded-direct, confounded-confounded,* and *other's-direct.* The degree of deterioration of RA varies considerably in these three cases and is almost imperceptable in the *confounded-direct* condition.

The third general relationship occurs in the three remaining cells and is characterized by a simultaneous deterioration of S^rM and RA. An S in one of these conditions receives poorer and poorer scores. He shows no signs of being able to return himself to an optimal set of reinforcing values and the response-magnitude has deteriorated considerably from preliminary training. This interesting type of S^rM-RA relationship is related exclusively neither to a particular row nor column. It appears to occur below the *diagonal* (from top right to bottom left). Recent additional study has demonstrated that this diagonal is, in fact, the boundary for this third type of relationship.

There are a number of interesting implications in these results for interpersonal adaptation and for the stability of a relationship. We have observed, for example, that in certain reinforcing relationships the interpersonal rewards may remain constant while physical features of the response are gradually modified. Such a reward structure would appear to result in a stable dyadic relationship, in the sense that the person will continue to interact in the absence of any restraints. However, social situations characterized by the third kind of relationship discussed above may be highly unstable because no apparent change in response will maintain a rewarding state of affairs for an individual. If there are no constraints on such a dyad, as there are in this experimental setup, the relationship would probably terminate rapidly. These inferences about the stability of the interpersonal relationship are testable by extensions of the experimental procedure where an S is permitted to choose among knobs (or response modes) for his response on each trial. Different reinforcement schedules would be assigned to the different knobs and the "sociometric" choices of the S noted as a function of these schedules.

The results of this study also have a number of implications for program-

ming reinforcing events to a single S so as to simulate social reinforcement, and for generalizing these results to interpersonal action. It is tempting, methodologically, to reduce interpersonal events to parameters which can be programmed for an individual. The crux of such reductionism, at least in the context of the present study, is the ability to specify, for each reward system, the S^rM values to an individual over a series of trials. It is a considerable oversimplication to dismiss the social component of S^rM as random noise, i.e., as containing a component random with respect to the S's response. Such a suggestion requires at minimum the further specification of parametric values of this noise. It would appear that the parameters of such a random component are difficult to specify and, further, that these parameters change considerably over time. Nor is it known in advance in what direction these parameters will change, i.e., whether the S^rM becomes more noisy or less noisy, or that these changes will be linear.

An examination of the reinforcement conditions in this experiment may clarify these points. First, the reinforcement linkages suggest two ways in which S^rM values may be programmed to an S over a series of trials:

Method 1. A predetermined series of S^rM values are given to an S regardless of what he does (represented in *others-to-S-direct-to-P*).

Method 2. A predetermined degree of contingency between the response of an S and his S^rM value is imposed (represented in *confounded-to-S-direct-to-P*).[7]

In the examples cited for both methods, the behavior of P is a well-learned response maintained under "direct" information and, hence, it is readily obtained and used as a component of S^rM. In fact, the parameters of P's behavior under *any* nonsocial conditions are readily obtained empirically. The point is that nonreciprocal social linkages can be "reduced," at least to a large extent, to individual psychological experiments.

How does one proceed with the programming of S^rM values in reciprocal linkages? Consider, for example, the programming of S^rM values in the reciprocal linkage, *confounded-to-S-confounded-to-P*. The S^rM curve for this condition in Figure 2 appears to fit well to a horizontal straight line. One could hypothesize that even if this is a reciprocal social linkage, the important determinant of RA is the series of S^rM values. This implies programming by Method 1 above (a predetermined set of values) to represent this condition.

[7] Note that the variations in reinforcement conditions used in this study correspond somewhat to distinctions betwen experimenter-controlled, subject-controlled, and experimenter-subject-controlled events made by Bush and Mosteller (1) and to similar distinctions of contingent and noncontingent events (e.g., 11) provided the experimenter is regarded as "partner."

Note, however, that the S^rM values in the *confounded-to-S-confounded-to-P-condition* appear to be quite similar to those in *others-to-S-direct-to-P-condition* and that the latter *is* an application of Method 1. However, the RA curves in these two conditions show quite different rates of deterioration. Hence, the hypothesis that the important determinant in reciprocal linkages is the value of S^rM is shown to be empirically false. Suppose, then, that in reciprocal linkages, one uses Method 2 above (a predetermined degree of RA-S^rM contingency) to program S^rM values to represent the *confounded-to-S-confound-to-P-condition*. This procedure would leave, unspecified, a method for programming P's behavior in the S^rM. Obviously, the response information required to reduce P to a programmed variable in reciprocal linkages is the very response data which are sought by such an experiment unless one is willing to guess at the constantly changing degree of correlation between S's and P's behavior! Perhaps a third method is appropriate, i.e., one which satisfies, simultaneously, the requirements in both Methods 1 and 2 above. The third method would require that S^rM values be contingent, to a specified degree, upon the S's behavior, and yet these values are not permitted to exceed a range of magnitudes. This range may, of course, change with trials as is true in three of the four linkages. Note, nevertheless, that it is still an empirical matter as to whether RA results under the third method of programming would resemble those obtained where the added ingredient of actual social reciprocal linkage is present. Note, also, that in order to apply findings derived by the programming (or individually oriented) approach to social events, it is necessary to know what happens to the S^rM in these social events and that such knowledge is inescapably obtained only by simulating these social events or by a well-tested theory of interpersonal learning.

SUMMARY

The present study has manipulated the reinforcing linkages between two persons, after each had first learned an operant in isolation. The experiment concerns the maintenance of the operant as a function of the reinforcement linkages between a Subject, S, and his Partner, P.

Nine reinforcement linkage relationships (as the independent variable) were generated by assigning the S to one of three conditions and the P to one of three conditions and representing all combinations (3×3) in the study. The three conditions are the extremes and middle point of a continuum which specifies the proportion of person A's response relative to person B's response represented in the reinforcement magnitude to person A.

A total of 54 persons was used in the study. They were paired randomly with one designated as the S and the other as his P. All persons were first

given 25 individual training trials on a simple knob-turning task followed by 50 experimental trials in one of the 9 reinforcement linkages.

Two dependent variables were analyzed for each of the 50 experimental trials: (a) response accuracy of the S; (b) reinforcement stimulus magnitude delivered to the S (or response "function"). Analyses revealed a number of significant differences between conditions on each of these measures. A S assigned to an experimental condition in which he continued to receive reinforcement based upon his responses only showed best maintenance of response accuracy while a S assigned to a condition in which he received reinforcement based upon P's responses only showed greatest response deterioration. The analyses of the second measure demonstrated effects somewhat different from those obtained for response accuracy. Deterioration of S's response function (second measure) depends upon both S's and P's conditions and upon an interaction of these conditions.

The covariation between the two measures was not sufficiently close, therefore, to provide either a simple summary of both dependent variables or to permit an "explanation" of one variable in terms of the other. Indeed, an examination of each of the nine conditions pointed up a number of interesting differences between the measures. These differences suggested further experimental analyses and were interpreted as indicative of important methodological principles in an experimental analysis of interpersonal phenomena.

Manuscript received: July 28, 1958
Revised manuscript received: January 26, 1959
Seymour Rosenberg
Bureau of Child Research
University of Kansas
Lawrence, Kansas

REFERENCES

1. Bush, R. R., and F. Mosteller, *Stochastic Models for Learning*, New York: Wiley, 1955.
2. Dollard, J., and N. E. Miller, *Personality and Psychotherapy*, New York: McGraw-Hill, 1950.
3. Gibson, J. J., "The Implications of Learning Theory for Social Psychology" in J. G. Miller, ed., *Experiments in Social Process*, New York: McGraw-Hill, 1950.
4. Gragg, D. B., and H. J. Douglass, *Conversion Tables for Selected Airman Classification Battery Scores and Comparable Scores on Other Selected Service and Civilian Tests*, Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center, June, 1956. (*Development Report* AFPTRC-TN-56–89.)
5. Grant, D. A., "Analysis-of-variance Tests in the Analysis and Comparision of Curves," *Psychological Bulletin*, 1956, 53, 141–154.
6. Hall, R. L., "Group Performance under Feedback that Confounds Responses of Group Members," *Sociometry*, 1957, 20, 297–305.
7. Hays, D. G., and R. Bush, "A Study of Group Action," *American Sociological Review*, 1954, 19, 693–701.

8. Keller, F. S., and W. N. Schoenfeld, *Principles of Psychology,* New York: Appleton-Century-Crofts, 1950.
9. Krasner, L., "Studies in the Conditioning of Verbal Behavior," *Psychological Bulletin,* 1958, 55, 148–170.
10. Miller, N. E., and J. Dollard, *Social Learning and Imitation,* New Haven: Yale University Press, 1941.
11. Neimark, D., "Effects of Type of Nonreinforcement and Number of Alternative Responses in Two Verbal Conditioning Situations," *Journal of Experimental Psychology,* 1956, 52, 209–220.
12. Rosenberg, S., and R. L. Hall, "The Effects of Different Social Feedback Conditions upon Performance in Dyadic Teams," *Journal of Abnormal Social Psychology,* 1958, 57, 271–277.
13. Sears, R. R., "A Theoretical Framework for Personality and Social Behavior," *American Psychologist,* 1951, 6, 476–483.
14. Sidowski, J. B., L. B. Wyckoff, and L. Tabory, "The Influence of Reinforcement and Punishment in a Minimal Social Situation," *Journal of Abnormal Social Psychology,* 1956, 52, 115–119.
15. Sidowski, J. B., "Reward and Punishment in a Minimal Social Situation," *Journal of Experimental Psychology,* 1957, 54, 318–326.
16. Skinner, B. F., *Science and Human Behavior,* New York: Macmillan, 1953.

APPENDIX

Summary of Statistical Analyses of the "Replication" Data

	df of Each Analysis	Response Accuracy		Stimulus Magnitude	
		Mean Sq.	F	Mean Sq.	F
1. Rows (Ss' conds.)	2	701,167	29.27‡	227,281	6.99†
2. Columns (P's conds.)	2	103,820	4.33*	249,903	7.69†
3. Rows × columns	4	47,364	1.98	120,531	3.71*
4. Within cells	18	23,955		32,507	
5. Trials	49	3,404		2,097	
6. Rows × trials	98	1,953	1.75†	1,554	1.16
7. Columns × trials	98	1,476	1.32*	1,658	1.24
8. Rows × columns × trials	196	1,184	1.06	1,524	1.14
9. Within cells × trials	882	1,115		1,334	

* = $P < .05$.
† = $P < .01$.
‡ = $P < .001$.

NOTE: Error terms employed in the computation of the *F*'s above are the same as those used in Table 1. Orthogonal polynomial breakdowns were not performed with the "replication" data.

A Description of Strengthening and Weakening Members of a Group [1]

IAN C. ROSS, *Bell Telephone Laboratories*
FRANK HARARY, *Princeton University*

This paper is one of a continuing series in which matrix analysis is used to identify various aspects of group structure. These include redundancies (6), complete cycles (3), liaison persons (7), and cliques (4). The structural concepts are developed elsewhere (1, 2).

THE PROBLEM

There are groups in which some members are seen as exercising a disruptive or divisive influence while other individuals appear to help hold the group together. Although there are a number of ways in which these weakening and strengthening effects of group members may be conceptualized, we shall confine our discussion to one that is suggested by and can be coordinated to various kinds of connectedness of "directed graphs" or "digraphs."

A digraph becomes a mathematical model for the structure of a group of people when its points represent the members of the group and the (directed) lines the relationships between pairs of members. The group structure may be built from many different kinds of relationships, such as sociometric choice, communication, or power. Hence it may be useful to have a way of characterizing, in digraph terms, the way in which individual members contribute to the connectedness of the group.

In order to give precise meanings to these concepts, we coordinate the structure of a group to a digraph. Then the kinds of connectedness a group may possess correspond to the four kinds of connectedness a digraph may have. A *digraph* is a finite set of *points* $A, B, C, \ldots$ and a subset of the *(directed) lines* $A \to B$, $B \to A$, $A \to C$, $C \to A$, $B \to C$, $C \to B$, ... joining pairs of distinct points. If there is a set of directed lines of the form $A \to D$, $D \to C$, $C \to B$ there is said to be a *directed path* from A to B. The *length* of this path is equal to the number of lines in it and the *distance* between A and B is the length of a shortest path from A to B. Although the digraph model is capable of describing several different kinds of bonds or lines at the same time, we shall confine ourselves to the simplest case in which the existence of lines implies the same characteristic of the relationship between each pair of

[1] This work was begun while the authors were at the Research Center for Group Dynamics, University of Michigan.

joined points. This restriction is not serious since the meaning coordinated to the lines may be a summary of many observations as, for example, an observer's judgment that person A chooses B. The situations in which one person chooses another may be quite diverse but by summarizing at the conceptual level, a digraph with just one type of lines is adequate. This example also serves to show that directed lines are more useful than symmetric lines since it is not necessary that B reciprocate by choosing A. He may or may not, but both possibilities can be handled. If A chooses B while B chooses C, then A is indirectly choosing C via B. This is one meaning of a two-step connection in digraphs. When other relationships are coordinated to the lines, the meaning of the multistep path may be more or less obvious but mathematically no different. If, for example, the lines are coordinated to "can give information to" then a multistep path indicates a possible channel for communication.

A digraph is strongly connected (or *strong*) if for every pair of distinct points, A and B, there exists a directed path from A to B *and* one from B to A. A digraph is unilaterally connected (or *unilateral*) if for every pair of points, A and B, there is a directed path from A to B or one from B to A. A digraph is *disconnected* if the points can be divided into two sets with no line joining any point in one set with a point in the other set. A digraph is weakly connected (or *weak*) if it is not disconnected. These connectedness definitions are *inclusive* since every strong digraph is unilateral and every unilateral digraph is weak.

In order to distinguish between groups on the basis of the kind of connectedness, we require exclusive connectedness categories. These may be obtained by modifying the inclusive definition as follows. Let U_3 be the collection of all strong digraphs. We define U_2 as the set of all unilateral digraphs that are not strong. Similarly we define U_1 as the set of all weak digraphs that are not unilateral. Finally, U_0 is the collection of all disconnected digraphs. Clearly, any digraph belongs to exactly one of the categories U_3, U_2, U_1, or U_0.

In order to evaluate the effect of a point on the connectedness of its digraph, we require a precise definition of the removal of a point of a digraph. If D is a digraph and A is a point in it, than $D-A$ is the digraph obtained from D by deleting the point A and all lines which are either directed toward A or away from A. We say A is a point of type P_{ij} if the digraph D (with A present) is in class U_i but $D-A$ (with A absent) is in class U_j. Since the four categories U_3, U_2, U_1, and U_0 are numbered in accordance with the convention that the higher the subscript the stronger the kind of connectedness of the digraph, we may utilize this convention to describe and characterize which points are strengthening. Thus a point of a digraph is a *strengthening*

point if and only if it is of type P_{ij} such that $i>j$; it is a *weakening point* if $i<j$ and the point is called *neutral* if $i=j$. By *strengthening and weakening group members* we mean those individuals coordinated with strengthening and weakening points of the digraph that represents the structure of the group. Similarly, if a point is of type P_{ij}, we speak of the corresponding individual as an *(i, j) member*.

When we contrast D and $D-A$ we assume that the lines between pairs of distinct points exist independently. In real groups the relationships between members may be dependent upon one another in ways that are not fully accounted for by our model. However, we believe that the structural changes reflected by the digraph model is a useful first approximation to the results of a member's departure from a group.

CONNECTEDNESS CRITERIA

We present a logical sequence of theorems that characterize strengthening and weakening group members. We first show that there are no (1, 3) members in a group. We then find it convenient to introduce the concept of the "reachability matrix" of a group. This matrix, R, is useful for the expression of matrix conditions that characterize the inclusive connectedness categories (i.e., strong, unilateral, weak, and disconnected). A straightforward modification of these theorems leads to a description of the exclusive connectedness categories U_3, U_2, U_1, U_0. We then obtain conditions for each of the *(i, j)* types. A particularly interesting result is that weakening members may be identified from direct examination of the reachability matrix.

To illustrate all the possible types of *(i, j)* members that can occur, we display the following digraphs in which the ordered pair written at each point indicates its type. Whenever a line between two points of any digraph of Figure 1 is drawn without any direction displayed, it stands for two directed lines, one in each direction.

Theorem 1. There are no (1, 3) members in any group, and all other *(i, j)* members can occur.

Proof. The validity of the second part of the theorem follows at once from the examples presented in Figure 1.

We assume that there exists a (1, 3) member A of group G, and derive a contradiction. Here G is in class U_1 and $G-A$ is strong. But when we join the point A to the group $G-A$ by at least one line, we see that the original group G must be unilateral (i.e., is in either class U_2 or U_3). This contradicts the hypothesis that G is in class U_1.

We now require the following definitions: Let $A_1, A_2, \ldots, A_n$ be the n members of the group. Let the matrix M of the group relationship be defined as follows: If A_i has the relationship to A_j, the i, j entry of M equals 1;

otherwise it is 0. We find it convenient to adopt the convention that all the diagonal entries are 1. The relationship is *not* assumed to be symmetric, but may be.

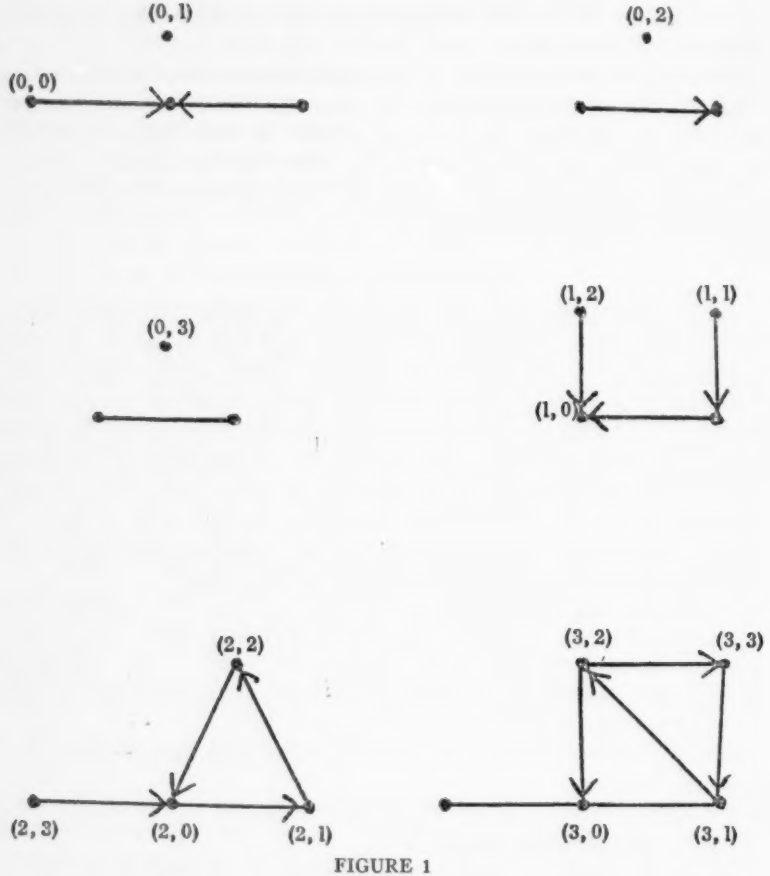


FIGURE 1

Let R denote the *reachability matrix* of the digraph or of the group with respect to this relation. Then the i, j element of R is taken as 1 if there exists a directed path from A_i to A_j and is 0 otherwise.

The *diameter, d,* of a digraph D is the greatest distance between any two points of D.

A *loop* is a directed line which begins and ends at the same point. Our convention that all the diagonal elements of M are 1 is equivalent to the existence of a loop at every point.

A *Boolian matrix* is one in which all the elements are either 1 or 0. The usual addition and multiplication operations of matrices are applied to Boolian matrices by means of the rule $1+1=1$. Thus in Boolian terms, any positive integer obtained by ordinary matrix operations is replaced by a 1. If the sum or product is 0, it is entered in the matrix as 0. In this paper all matrices are Boolian, and all the operations are performed as described above.

A redundant chain from B_1 to B_k is a sequence of lines of the form $B_1 \rightarrow B_2$, $B_2 \rightarrow B_3$, ..., $B_{k-1} \rightarrow B_k$ where the points B_1 to B_k are not all distinct.

The following theorem gives a formula for the reachability matrix of a group in terms of its relationship matrix.

Theorem 2. $R=M^d$

Proof. By its constructive definition the matrix M gives all the (directed) paths of length 1 between distinct pairs of points. The matrix M contains all these paths and also a path of length 1 from each point to itself. Thus each point is reachable from itself. In addition, these loops permit the existence of redundant chains from A to B of any length whenever paths from A to B of shorter length occur. We now form the matrix M^2. In this matrix, the number 1 occurs in the i, j place if and only if there is either a path or redundant chain of length 2 from point i to point j.

Continuing in this manner we eventually obtain the matrix M^d. Since d is the diameter of the given digraph, there are no paths of length greater than d that connect points not already connected. The presence of loops at each point assures us that there will be a redundant chain of length d whenever there is a path of any shorter length. Therefore this last matrix M^d contains all reachability relationships that occur in the digraph. Hence, $R=M^d$.

Corollary 1. If k is any integer greater than d, then $M^k=R$.

Proof. We know that the diameter has been reached the first time two consecutive powers of M are equal. For the diameter d is the smallest number m such that $M^m=M^m+1$. Hence for any integer k greater than d, the matrix M^k is equal to M^d which we have already seen equals R.

We now apply this corollary to get a *working* procedure for computing R. After finding M, calculate M^2, M^4, M^8, etc. As soon as two consecutive matrices of this sequence coincide, we have found R. This procedure usually calculates R in the fewest possible matrix operations.

Corollary 2. $R=(I-M)^{-1}$.

Proof. By a well-known matrix identity, $(I-M)^{-1}=I+M+M^2+\ldots+M^d+M^{d+1}\ldots$ But the terms $M^d+M^{d+1}+\ldots=M^d$ since we are using Boolian

addition. Also $I+M+M^2+\ldots+M^d=M^d$. Hence $(I-M)^{-1}=M^d=R$ by Theorem 2. We note that if N is the matrix obtained from M by having 0's on the diagonal, then $R=(-N)^{-1}$.

Inclusive Connectedness Categories

The next four theorems serve to characterize the inclusive connectedness categories. These are expressible in terms of the matrix V of the *universal relationship*. Specifically, V is the n by n matrix in which every entry is 1.

Theorem 3. The group G is strong if and only if $R=V$.

Proof. By definition, G is strong if and only if for each member-pair A_i and A_j there exists a path in both directions. However, this condition holds if and only if for all values of i and j, the i, j element of the matrix R is equal to 1. However, i and j are arbitrary positive integers between 1 and n. Therefore all the entries in R must be equal to 1 for the group to be strong, i.e., $R=V$. (This result was also found by Rosenblatt (5).)

The transpose M' of a matrix M is obtained by reflecting the entries of M about its diagonal, i.e., by interchanging the rows and columns of M. For example, R' is the matrix of "reverse reachability" in the sense that a 1 in the i, j place of R' means that A_i is reachable from A_j. The next theorem uses the matrix $R+R'$. In this matrix, a 1 in the i, j place means A_j is reachable from A_i, or A_i is reachable from A_j, or both (since the addition of R and R' is Boolian).

Theorem 4. The group G is unilateral if and only if $R+R'=V$.

Proof. Let G be unilateral and let A_i and A_j be any two of its members. Then by definition of unilateral, either A_j is reachable from A_i or A_i is reachable from A_j (or both). In the former case the i, j element of R is 1 while in the latter case the i, j element of the transposed matrix R' is 1. Therefore regardless of which of these alternatives hold, the i, j element of $R+R'$ is 1. But i and j are arbitrary integers between 1 and n. Hence, $R+R'=V$. On reversing each of these steps it follows that if $R+R'=V$ then G is unilateral.

In order to characterize weak groups, we introduce the digraph D obtained from D as follows: First, D^- is a supergraph of D, i.e., all points and lines of D are in D^-. Second, if the directed line $A\rightarrow B$ is in D, then both of the lines $A\rightarrow B$ and $B\rightarrow A$ are in D^-. This process may be called *symmetrizing* the digraph D. Let M^- and R^- be the relationship matrix and the reachability matrix of D^- respectively. Then $M^-=M+M'$.

Theorem 5. A group is weak if and only if $R^-=V$.

Proof. Obviously D is weak if and only if D^- is strong. Therefore, by Theorem 3, D is weak if and only if $R^-=V$.

Theorem 6. A group is disconnected if and only if $R^-\neq V$.

Proof. By definition, a group is disconnected if and only if it is not weak. Therefore, a criterion for it to be disconnected is, by Theorem 5, that $R^- \neq V$.

Exclusive connectedness categories.

We now combine the last four theorems to characterize the exclusive categories U_3, U_2, U_1, and U_0.

Theorem 6 gives a criterion for a group G to be in U_0. However, for G to be in U_1 the condition of Theorem 5 holds while that of Theorem 4 does not. Similarly, G is in U_2 if and only if Theorem 4 holds while Theorem 3 does not. Finally G is in U_3 if and only if Theorem 3 holds. These observations are summarized in Table 1 which lists the exclusive connectedness categories in one column and the respective criteria in the other column.

TABLE 1

Exclusive Class	Criterion
U_3	$R=V$
U_2	$R+R'=V, R\neq V$
U_1	$R^-=V, R+R'\neq V$
U_0	$R^-\neq V$

WEAKENING MEMBERS

The weakening members of a group can be described further. All the possible kinds of weakening members are:

A. The $(0, j)$ members for $j=1, 2,$ or 3

B. The $(1, 2)$ members

C. The $(2, 3)$ members

This follows from Theorem 1 which asserts that there are no $(1, 3)$ members.

There is independent interest in identifying the isolates in a group.

Theorem 7. The member A_i is an isolate if and only if the only nonzero element in the ith row and the ith column of R is the i, i (or diagonal) entry.

Proof. Obvious.

A *symmetric matrix* is one which remains unchanged when its rows and columns are interchanged. It is well known that for any square matrix X, the matrix $X+X'$ is symmetric. Hence in particular $R+R'$ is symmetric.

Theorem 8. (Characterization of all weakening members)

I. A_i is a $(0, j)$ member for $j=1, 2,$ or 3 if and only if A_i is an isolate and $G-A_i$ is in U_j.

II. A_i is a $(1, 2)$ member if and only if $R+R'$ has at least one zero and all its zeros occur in the ith row and in the ith column.

III. A_i is a (2, 3) member if and only if every element in R is a 1 except in the ith row (or column) of R, where all but the diagonal element are 0.

Proof of I. If A_i is an isolate and $G-A_i$ is in U_j then A_i is a $(0, j)$ member by definition.

To prove the converse, let A_i be a $(0, j)$ member for $j>0$ and assume A_i is not an isolate. Then $G-A_i$ is still disconnected. Hence A_i is a $(0, 0)$ member which is a contradiction.

Procedurally, $(0, j)$ members can be identified by considering the condition met by the remainder of the matrices R, $R+R'$, and R^- after the row and column belonging to an isolate are deleted.

Proof of II. If $R+R'$ has at least one zero and all its zeros occur in the ith row and in the ith column, then by Table 1, G is in U_1 and $G-A_i$ is in U_2.

To prove the converse, let A_i be a (1, 2) member of G. Then G is in U_1 so that $R+R'$ must have at least one 0. Assume that there is a 0 which is neither in the ith row nor in the ith column of $R+R'$. If A_i is eliminated from $R+R'$, all 0's not in the ith row or column would remain since the elimination of a member and its bonds cannot add connections to members of G. Therefore $G-A_1$ would be in U_1, a contradiction.

Proof of III. Again the criteria of Table 1 show that under the stated conditions A_i is a (2, 3) member.

Conversely, let A_i be a given (2, 3) member. Then either A_i can reach all other members but no other members can reach A_i or vice versa. In the first case all the elements in the ith row are 1 and all the nondiagonal elements in the ith column are 0, while every element of R not in the ith row or column is a 1. For the other case, we interchange the words "row" and "column."

Corollary I. The group consisting of exactly two isolates has two (0, 3) members. Any other disconnected group has at most one weakening member.

Corollary II. A U_1 group has at most two (1, 2) members.

Corollary III. A U_2 group has at most two (2, 3) members.

Proof. These limitations on the number of weakening members follow readily from the characterizations of the preceding theorem.

The following is probably the most startling result in this entire paper. It must be kept in mind that weakening members are defined in terms of the kinds of connectedness introduced here.

Theorem 9. Any group has at most two weakening members.

Proof. This is an immediate consequence of the three parts of the above corollary since the possibilities there listed are exhaustive and mutually exclusive.

Unfortunately there do not appear to be analogous theorems describing the strengthening members of a group. However we can still identify these

strengthening members by using the results of Table 1 on the given group G and on $G-A$. In particular, the strengthening members in classes $(i, 0)$ for $p=1, 2$, or 3 correspond to the liaison persons (7) of the symmetrized group.

Manuscript received: April 18, 1958
Revised manuscript received: December 11, 1958
Ian C. Ross
Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey

REFERENCES

1. Harary, F., and R. Z. Norman, *Graph Theory as a Mathematical Model in Social Science*, Ann Arbor, Mich.: Institute for Social Research, 1953.
2. Harary, F., R. Z. Norman, and D. Cartwright, *Introduction to Digraph Theory for Social Scientists*. In process of publication.
3. Harary, F., and I. C. Ross, "The Number of Complete Cycles in a Communication Network," *Journal of Social Psychology*, 1953, 40, 329–332.
4. Harary, F., and I. C. Ross, "A Procedure for Clique Detention using the Group Matrix," *Sociometry*, 1957, 20, 205–215.
5. Rosenblatt, D., "On Linear Models and the Graphs of Minkowski—Leontief Matrices," *Econometrica*, 1957, 25, 325–338.
6. Ross, I. C., and F. Harary, "On the Determination of Redundancies in Sociometric Chains," *Psychometrika*, 1952, 17, 195–208.
7. Ross, I. C., and F. Harary, "Identification of the Liaison Persons of an Organization using the Structure Matrix," *Management Science*, 1955, 1, 251–258.

A Broad Empirical Approach to Personality: Personality and Motivation Structure and Measurement [1]

EDGAR F. BORGATTA, *Russell Sage Foundation*

Raymond B. Cattell has been a prolific and tireless worker in the areas of intelligence, personality, and group analysis, and this volume represents an occasion for reviewing his general approach, particularly as relevant to social psychology. The review presented here is an attempt to describe in a very brief excursion some of the topics with which he deals and the position he takes on certain issues. It does not, however, attempt to contrast Cattell's work with that of others, nor is it a critique of the findings reported. Either of these tasks could constitute a volume in itself.

Empirical research is a keynote in identifying Cattell and his position in psychology, for he does not operate with what might easily be termed the "grand theories." Rather, he is rarely more than a few steps beyond a statement of a set of empirically derived generalizations. The mainstay of Cattell's approach seems to be that of identifying the important and relevant variables in the common intuitive way, attempting to construct measures inclusive of these, and then, having accumulated many measures, sorting these out according to the common expectations of the language and the field and also according to mechanical routine procedure. The emphasis is in creating an opportunity for finding order rather than imposing it by preconceptions that might limit either the scope of inclusiveness or of analysis.

Cattell introduces his volume through an essay (Chapter 1) that assesses the current status of psychology in regard to personality and measurement, and more generally as a science. His phraseology must sometimes wave a red flag before the eyes of some more self-conscious "professional" psychologists. For example, "Unfortunately, psychology, during the first fifty years of the twentieth century, has been largely a farrago of popular, philosophical, verbal punditry. The theories of many educational and learning theorists, psychoanalysts, and other followers of applied psychology, for example, are devoid both of measurement and the possibility of measurement" (p. 3). While the caricature may be correct, polite discourse requires that this be left unsaid or said more subtly. Note is taken of this occasional tendency in Cattell's writing since it would be extremely unfortunate if any reader would use it as a justification for not giving Cattell's work serious consideration. Cattell, in a similar manner, speaks of the *factor analyst,* thus fostering this tendency

[1] Cattell, R. B., *Personality and Motivation Structure and Measurement,* Yonkers, N.Y.: World Book Company, 1957.

to identify persons by the techniques employed rather than on the merits of the findings and research reported. This might not be bad if persons were equally willing to think of Psychologist X as a *"t" tester,* and Psychologist Y as an *analysis-of-variance-ist.* A third thing that Cattell does that might serve as a distraction is that in his criticism of particular work and trends, he at times criticizes abuses associated with a particular approach rather than the approach itself (e.g., in the treatment of cluster analysis, pp. 20ff.).

Cattell's approach may be characterized as one of multiple variable classification and the development of a system relating sets of classes within the taxonomy. He emphasizes the rudimentary state of knowledge in psychology and the social sciences today and decries the premature stabs at experimental manipulation in the absence of proper variable definition. He is not anti-experimentalist, but he does emphasize the descriptive requirements in science as primary. For example, "Unfortunately, the archaic views of what constitutes experimental psychology now prevailing in certain universities have caused manipulative experimenters, especially in perception and learning, to waste time for a generation experimenting with single variables when they should have been experimenting with the relations of proven unitary concepts, i.e., with factors established by preliminary multivariate studies. Although the present book strains at the leash to reach manipulative experiment on personality, and at several points reaches that stage, it has to be concerned in the main with the multivariate evidence for the existence and nature of the chief stable, functionally unitary personality and motivation patterns" (p. 27, 28).

The suggestion that Cattell is atheoretical or untheoretical in his approach is misplaced. Rather, and Cattell identifies this himself, he is not concerned with what might be called the fashionable theories. In considering them as such, for example, he says: "Thus we may find that interest switches from year to year—now to 'authoritarian traits,' now to 'achievement,' and so to 'anxiety,' 'co-satiation behavior,' or whatever was last bruited. And in historical retrospect we have, in the journals, vintage years for rigidity, problem-solving ability, etc., etc." (p. 49). In other words, Cattell steers away from and rejects the involvement in single overriding variables as a basis of understanding in psychology and the social sciences. On the other hand, he explicitly takes as a point of departure in developing the relevance of items and variables for consideration *all* the prior work. This may be a grandiose and ambitious task, but it will have to be judged historically rather than on the basis of personal predilections.

Assuming, thus, that if a variable at some point has been judged to have relevance to the proper study of psychology and the social sciences, then it is within the scope of Cattell's interest. Initially he is not concerned with the

classification of these but in the classification of the means of observing the variables. His model for classification is as follows: (a) First, there is the life record which constitutes situational or criterion variables. If there is an ordering of behavior that is to be predicted and understood, presumably this is the behavior that is ultimately in question. (b) The second class of observations are those ascertained through questionnaire techniques, and these may refer to introspective or interior variables, and also to objective type variables that constitute the third class. (c) The third class are the objective type variables that may in some cases be obtained by questionnaires but are not self-evaluative as such. The objective test variables also include the other classes of direct and indirect measures.

Cattell organizes the variables within each of these classes of observation. The organization, thus, is based (following the factor analytic model) on the isolation of orthogonal categories so far as they exist. Where clusters do not exist that are orthogonal, the task is the identification of the relationships of the clusters for which some independent or orthogonal content has been isolated. A second stage of analysis centers on the examination of factors *through* the three media of observation. In part this is an orientation of reductionistic interpretation, although in his handling of it Cattell generally presents this as a regression problem.

A substantial segment of Cattell's volume is devoted to the presentation of the factors identified according to the three classes of observations. Additional attention is given to the problem of prediction and specification, and this necessarily is interwound with a good deal of methodological problematics. The presentation in these contexts goes well beyond Cattell's own work and he does draw from other workers in the field, including Guilford, Thurstone, Eysenck, and many others. He is cognizant of the differences in personality and behavior patterns according to age, but he does not emphasize the exploration of developmental questions as such here. If his data are already complex with the many factors involved, tracing their development certainly would be even a more monumental task. It should be recognized, however, that some of the test data that enters into the specification of factors inevitably has a temporal reference, so that developmental questions are at least partially implicated even in the analysis at the given level.

Although the evidence in this volume indicates he has not as yet reached the point where his cumulated empirical researches really encompass some of the aspects in the social area, Cattell is certainly a broadly based social psychologist. His Chapter 10, titled "Complex Functions and Types in Personality and Cultural Interaction," presents in a fairly concise way much of his work at the "social" level dealing with the problems of cultural factors,

complex situations, and indeed with the problem of prediction across such discrete social phenomena as roles. Admittedly, here he is entering a field rather than working in it. His awareness of the problems is acute, however. For example, "The fact that there are 'socially prescribed limits,' e.g., that even the kindest policeman cannot omit reporting *some* traffic offenses, indicates from the beginning that inter-role variance in behavior is likely to exceed personality variance within role" (pp. 422–423). In his analysis he defines his concern as primarily with ". . . the *role action pattern,* i.e., what a person actually does in a role" (p. 423). He does point to additional questions that arise. "For example, by an r_p between the centroid profile for the type and the individual we could give meaning to *how much* a person has stepped into a role. A further question concerns the relation of the effectiveness of a person in a role (How good a postman or a father is he?) with the degree to which he has adopted the role" (p. 425). The problem of dealing with the social reality in this context is indicated in the following position: "Verbally this corresponds to saying that when Smith puts on his uniform he 'perceives situations in a new way' or that 'he becomes a different person.' The principle of parsimony and general psychological knowledge dissuade us from making him a new person every time he steps into a role and favor the interpretation that the situations are differently perceived, i.e., that they have a new psychological meaning for all people in the role, requiring new situational indices" (pp. 425–426). Possibly the ultimate interest of Cattell as a social psychologist is best expressed in his theoretical formulation of the space in which he does research, illustrated by the well-known diagram on page 494, with three directions for test standardization: variables, persons, and occasions.

Interest in the specific content of Cattell's researches will center on at least two levels. In the first, the specification of the test analyses in terms of the three media of observation will be important. In these cases the effort in analysis is the presentation of the specific identification of the factor variables, and this is suggestive, in terms of hypotheses both in the abstract empirical sense of Cattell's own analysis and also with reference to other theoretical systems.

The more complex theory of Cattell, more closely associated with the grand synthesis of personality and the understanding of individual behavior, is found in Part IV of the volume, "The Measurement of Attitudes, Motivation, and Conflict Adjustment," and Part V, "Personality Change: Measurement of Anxiety, Fatigue, and Psycho-Physiological States." This is in many ways the most interesting and speculative side of Cattell, but also is necessarily the most tenuous aspect of the empirically based theory. Cattell deals with the

applied aspects of psychology and the relationship of his work to these in a section in the Part titled "Resultant New Test Techniques in Clinical, Educational, and Industrial Psychology."

Cattell's volume is a large one both in its 900 pages and in content. As a cumulation of a considerable experience in textbooks and articles, it is well organized and includes many useful devices, including appendices that are themselves rich outlines of some aspects of the material reported.

This review has attempted to present Cattell in terms of the contributions derived from his work. It has avoided laboring the specific criticisms that might be made of this or that particular statement or assumption, but has emphasized the scope and depth of Cattell's interest and presentation. It is possible to quarrel with Cattell on many issues, such as the meaningfulness and utility of second order factors, but it is not possible to ignore his substantive contribution. On the one hand, this is of such proportions that some members of the profession would prefer to ignore it than to tackle it. On the other hand, his broadly based multivariate approach becomes embarrassing to those who would like to make a lifetime occupation of examining the relationship of X_1 to X_2 when the data are already available for the relationships of X_1. In fact, the concern with studies of the relationship of X_1 to X_2 has dominated the scene sufficiently to lead Cattell (1) to indicate that integrative surveys themselves frequently ignored multivariate studies. This volume represents a tremendous amount of empirical research with a fair amount of digestion. It can be ignored only to the extent that broadly based systematic descriptive empirical research can be ignored. If it is beyond the clinician and demanding on the researcher, this is a function as much of the growing maturity and complexity of the field as of the shortcomings that Cattell, like any of us, may have.

Manuscript received: February 5, 1959

Edgar F. Borgatta
Russell Sage Foundation
505 Park Avenue
New York 22, New York

REFERENCE

1. Cattell, R. B., "A Need for Alertness to Multivariate Experimental Findings in Integrative Surveys," *Psychological Bulletin*, 1958, 55, 253–256.

The Effect of Acquiescence Response-Set upon Relationships among the F Scale, Ethnocentrism, and Intelligence [1]

LOREN J. CHAPMAN, *University of Chicago*
DONALD T. CAMPBELL, *Northwestern University*

When attitude and personality measures employing the endorsement of prepared statements contain only items worded in the direction of one end of the dimension, the intended content of the statements may be confounded with very general tendencies of acquiescing to declarative statements. This was pointed out quite a few years ago (10, 16, 18), and more recently its importance has been demonstrated for the F scale from *The Authoritarian Personality* study (2, 3, 4, 5, 9, 13, 14, 15, 19). All items in the F scale are worded in the positive direction, so that agreement with the item contributes to a high F score. Several investigators have approached the problem by rewording F items so that disagreement contributes toward a high F score, and then correlating the reversed scale with the original. All these investigators have found such correlations to fall far short of the reliability of the scales, and from this and other evidence have inferred that acquiescence response set is present in a strong degree, necessitating a reinterpretation of results previously obtained with the F scale.

Christie *et al.* (8) have reexamined these results and have argued that some of the failure to find substantial correlations in the content direction was due to the use of inappropriate item reversals. They have provided a new set and have claimed correlations higher than those found in the previous efforts. Since their emphasis was upon what they viewed as the relative success of their reversals where others had failed, they answer the question "Can the F scale be reversed?" with an emphatic "Yes." This emphasis may mislead the casual reader into believing that the Christie reversals are free of acquiescence bias, and that they demonstrate the original F scale to be free of such bias also. This is far from the case, as has been pointed out by Chapman and Bock (5), and will be discussed rather briefly in the present paper. The focus upon the relative success of one reversal vs. another takes attention away from the central problem: now that some degree of acquiescence bias has been identified in the reliable variance of F-scale scores, how are we to reinterpret the well-established correlations between it and other variables? Since the concept of response-set emerged in the context of studies of inter-

[1] This study was supported in part by the Graduate School, Northwestern University, and in part by the Social Science Research Committee of the University of Chicago.

relationships among tests, it is strange that so many of the F-scale response-set studies have been devoid of any emphasis on external correlations. This study examines the effect of direction of item wording upon the well-established correlations between the F scale and ethnocentrism, and between the F scale and intelligence. In addition, the role of acquiescence bias is examined for the California E scale (1) both in terms of the correlation between sets of items of reversed direction of wording and in terms of relationships with the F scale and intelligence.

METHOD

For each item of the standard forms of the two tests a reversal was prepared. Examples of the originals and their reversals are as follows: [2] Ethnocentrism scale: "Negroes have their rights, but it is best to keep them in their own districts and schools and to prevent too much contact with the whites." "It is bad to keep Negroes in their own districts and schools, and bad to prevent their having contact with whites." "I can hardly imagine myself marrying a Jew." "I can easily imagine myself marrying a Jew." F scale: "No weakness or difficulty can hold us back if we have enough will power." "It takes more than will power to conquer most obstacles." "When a person has a problem or worry, it is best for him not to think about it but to keep busy with more cheerful things." "When a person has a problem or worry he should try to think it through rather than try to forget it by keeping busy with more cheerful things." (Other examples of reversed F-scale items are presented in the item analysis in the report of results that follows.)

From the original and reversed items, two versions of each scale were prepared, each having half positive and half negative items, each having but one member of a given content-reversal pair. On a subsequent testing period one week later each respondent received the alternate version of each scale so that in the end each respondent had received all items in both wordings. For both E and F measures, each respondent was given a separate score for the positive items and for the negative items, as well as a total composite score. All items were scored in the same content direction: e.g., a person received one point for agreeing with a positive item or for disagreeing with a negative one. The items were all presented in a true-false answer format to eliminate any extremity of emphasis response set (10), although this set was not itself under investigation.

The respondents were 66 male and 118 female students in introductory psychology classes at a Midwest university, spring, 1956. From the uni-

[2] Full details on the rewordings are available from the authors. F-scale reversals were for the most part borrowed or adapted from those previously used by Bass (2) and Chapman and Campbell (4).

versity's admissions records it was possible to get intelligence test scores for many of the respondents. For 86 of the 184, O.S.P.E. percentile ranks were available. For another 70 a sum of the verbal plus mathematical scores on the College Boards was employed. For purposes appropriate to the intent of the F and E scale measures, Jewish subjects were eliminated on the basis of a religion question asked at the end of the second testing. Negro and Oriental students were eliminated by noting names during the testing period.

ANALYSIS, RESULTS, AND DISCUSSION

The Criteria of Evaluation

Two main criteria are available in the data which bear upon the question of acquiescence bias or upon the more general question of the effect of direction of item wording on the construct validity of the tests.

TABLE 1

Reliabilities and Intratest Correlations

	N	Reliability of Positive Scale	Reliability of Negative Scale	Reliability of Combined Scale	Positive-negative Correlation	Set-free Positive-negative Correlation
E	184	.80	.69	.89	.70	(.74)
F	184	.71	.42	.70	.29	(.54)

NOTE: For the E scale as here used there were 19 items in each half, for the F scale 30 items. The reliabilities are Kuder-Richardson Formula 20. The values are not corrected for attenuation.

If acquiescence set variance is large, then the correlation between the positive and negative items should be lower than the reliabilities of each taken separately. Table 1 presents these data. To make the comparison more direct, a hypothetical "set-free positive-negative" value has also been provided, indicating what the correlation between the positive and negative items would be if, over and above their differences in reliabilities, both sets measured the same thing. (This was computed by the formula for the correction for attenuation, entering the two obtained reliabilities and unity for the "true" relationship and solving for the correlation between positive and negative items. This amounts to the geometric mean of the two reliabilities.) Similarly, whereas normally doubling the length of a test, such as occurs in the combining of the positive and negative sets of items, increases reliability, the suppressor effects of the item-wording factor in this case may nullify the gain if acquiescence set is strong.

Second, if acquiescence set is present, then this unwanted source of systematic variance should augment the correlations between variables sharing

a common direction of wording, and depress that between variables of opposite wording. Table 2 present these data. The two italicized values should be higher than the two others in their tetrad if acquiescence variance is biasing the relationship. Because of the difference in reliabilities, one should perhaps better compare the sum of the two italicized values with the sum of the two others.

TABLE 2

Correlations among Measures

N		E_{pos}	E_{neg}	E_{tot}	F_{pos}	F_{neg}	F_{tot}
184	F_{pos}	*.57*	.39				
184	F_{neg}	.21	*.31*				
184	F_{tot}			.50			
86	O.S.P.E.	−.32	−.17	−.26	−.26	−.03	−.26
70	V. + M. College Boards	−.36	−.26	−.34	−.35	−.05	−.26

NOTE: The values are not corrected for attenuation.

For both measures under study, significant negative correlations with intelligence have been reported (e.g., 6). Messick and Fredericksen's findings (19) indicated that such correlations reflect primarily a negative correlation between intelligence and the tendency to find questionnaire items plausible. The two final rows in Table 2 present further evidence in this regard.

The Ethnocentrism Scale

In Table 1, the effect of acquiescence upon the E scale shows up as very slight. The obtained correlation between the positive and negative scales was .70, only slightly less than the ideal one of .74 (given the obtained reliabilities of the two parts). In Table 2, the effect of direction of item wording seems more marked—the correlation with the original F scale shifts 18 points, from .57 to .39, as a function of item wording. This difference was found by Hotelling's technique (22, p. 257) to be significant ($p<.001$). A corresponding shift is found for the F_{neg} items although this difference fell short of significance ($p<.07$). All correlations are in the direction expected, and all are significantly different from zero (for N=184, an r of .19 is significant at the .01 level, an r of .24 at the .001 level). With regard to the correlation with intelligence, it seems possibly enhanced by acquiescence set, being higher for E_{pos} than E_{neg} although not significantly so for either measure of intelligence. It is also apparent that response to E content, apart from acquiescence bias,

is significantly correlated with intelligence. In general, it seems probable that the correlations reported between the E scale and other variables are in major part a function of content.

The F Scale

Table 1 provides intratest evidence of acquiescence variance in the comparison of the expected positive-negative item correlation of .54 with the obtained one of .29. The fact that F_{tot} based on 60 items is no more reliable than F_{pos}, based on 30, is further evidence. Table 2 continues the picture, as we have already seen, with relationships with the E scale being augmented and depressed depending upon agreement or disagreement of direction of item wording. The effect of direction of item wording upon the correlation with intelligence is definite. F_{neg} shows almost no correlation; F_{pos} shows the typically found significant correlation. The difference between the correlations is significant ($p<.05$) for the College Boards and just short of significance ($p<.07$) for the O.S.P.E. (Note that these data imply that both F content and acquiescence are related to intelligence, summating in F_{pos}, canceling in F_{neg}. If it were only acquiescence, F_{neg} should correlate as highly as F_{pos}, but with a reverse sign.)

While the data give evidence of acquiescence bias, they also give evidence of an intended content factor successfully identified and represented in the F_{neg} items. The correlation of F_{neg} with F_{pos} (.29) and with E_{pos} and E_{neg} (.21 and .31) are all significant and in the expected content direction. The .29 value is in general higher than F-scale reversal correlations previously reported.

The F reversals used in the present study were in large part borrowed or adapted from these previously used by Bass (2) and Chapman and Campbell (4) and thus consist largely of items criticized for suitability by Christie *et al.* (8). Christie's comments are so effectively presented that the use of any item set other than his seems to require special justification. Christie's sample most similar to ours is his 152 students from a northeastern university. For this group, his F_{pos}–F_{neg} correlation based on ten items each is only .05 (when his correlation for attenuation is removed). Other samples, to be sure, do show higher values, but as Christie points out, the size of the correlation varies with test-takers as well as with item reversals. A more direct comparison is available: L. J. Bookbinder[3] recently administered 15 Christie reversals and 15 original F items to an introductory psychology class at the same Midwest university from which our sample came, obtaining an F_{pos}–F_{neg} correlation of .17. A comparable value for 15 items vs. 15 items from the first session

[3] Personal communication.

of the present study is .14, only slightly less. On statistical grounds, then, the difference in adequacy of the Christie reversals and the present ones is slight, if it exists at all. Christie's reversals are, however, sufficiently different that it would be worth while exploring them also in relation to E-scale reversals and intelligence.

There are two general grounds upon which the F scale might be expected to be more susceptible to acquiescence bias than the E scale. On the one hand, each individual F-scale item might be judged to be more ambiguous, to have more complex, abstract, intangible referents, to be more "difficult," and to be therefore more prone to acquiescence set. But even if this difference were not true, the F scale has an item-to-item heterogeneity of topic far greater than that of the E scale, and this would produce greater susceptibility to acquiescence. For the original forms, the average interitem r inferred from the reliability coefficient is .174 for E_{pos}, as compared with .075 for F_{pos}. Cronbach (11) has shown that a factor which plays a minor role in each item can play a major role in the variance of the total score, if it is the only factor which is repeated from item to item.

Ethnocentrism and Personality

The original construction of the F scale was in part a deliberate testing of hypotheses and in part a matter of blind empiricism. The empiricism is represented in the selection of F-scale items through several repeated item analyses against the E-scale total score. Such a process allows the retained items to drift away in unrecorded directions from the original content-universe specified by the theory. In a sense, every time an item is dropped, the construct or content-universe has been redefined. Thus, for example, the superstition items retained in the final form of the F scale are not a random sampling of superstition items in general, nor even of the superstition items of their original item pool, but are rather a very special subset, the appropriate label for which is unspecified. If the empirical criterion used in selection itself contains systematic impurities, the item selection process will be biased in the direction of items predictive of those impurities as well as predictive of the desired content variance. The E scale almost certainly contains some acquiescence variance in addition to content variance. Note that it was only for the E scale that the deliberate decision was made by the authors of *The Authoritarian Personality* to have all items worded in a single direction, and that in so doing they went directly against the explicit recommendation of their methodological source (20, p. 283). The possibility that the final F-scale items were, in part, selected as good predictors of acquiescence variance is indicated by the fact that in the original item pool there were four items written in the non-F direction, all of which were later dropped because of

unfavorable item analysis results rather than for reasons of theory. Actually some of the item validity indices indicate that these items correlated with E in the reverse of the direction expected, one with a value of a magnitude that would have led to its retention had it been in correct direction.

The finally selected F-scale items are thus ambiguous indicators of the personality and attitudinal correlates of ethnocentric ideology. And while the total score correlations with the reversed E scale and with E_{tot} give confirmation that the general content relations are in the direction hypothesized, the correlations are so low and the content of the F scale so heterogeneous that what holds as a general trend might well fail of confirmation for some of the component syndromes. Christie and Garcia (7) have well documented the internal heterogeneity of the F scale.

For this reason, it seemed desirable to perform an item analysis of all the F-scale items, both F_{pos} and F_{neg}, against the measure of ethnocentrism judged most free of acquiescence bias, namely E_{tot}, to see which syndromes and which items of the F scale retained their relationships and which did not. The item analysis was done in terms of the Flanagan tables of phi coefficients (21).

Of the 30 F_{pos} items, the values ranged from $-.02$ to $+.60$, 29 showing positive correlations with E_{tot}. Of the 30 F_{neg} items, values ranged from $-.27$ to $+.42$, 23 showing positive correlations. In 25 out of 30 instances, the F_{pos} item was better than its F_{neg} partner. On some items both members of the pair correlated well, lending confidence to an interpretation of item content. These pairs will be identified by presenting the F_{neg} version, followed by the F_{pos} and F_{neg} correlation values:

Many people who do not feel a great love, gratitude, and respect for their parents are nevertheless very fine people. (.60, .42)

Strict discipline, rugged determination, and the will to work and fight for family and country are *not* what a youth needs most. (.55, .42)

Homosexuals are not criminals and should not be severely punished. (.48, .33)

It is fallacious to claim that people can be divided into two distinct classes, the weak and the strong. (.27, .32)

Even though nowadays many different kinds of people move around and mix together a great deal, it is foolish to worry much about catching an infection or disease from them. (.27, .28)

Sex criminals such as those who rape and attack children should be treated as sick people and not severely punished. (.17, .36)

On the other hand, a number of items in both F_{pos} and F_{neg} seem unrelated to ethnocentrism as measured by E_{tot}, in this sample:

It is unnecessary and undesirable for a person to have complete faith in some supernatural power whose decisions he obeys without question. (—.02, .00)

Nobody ever learned anything really important through suffering. (.00, —.20)

When a person has a problem or worry, he should try to think it through rather than try to forget by keeping busy with cheerful things. (.05, —.12)

Wars and social troubles will *not* be ended someday by an earthquake or flood that destroys the whole world. (.09, —.08)

If people would discuss matters more before acting, everybody would be better off. (.05, —.05)

The item providing the lowest negative validity was this one:

Nowadays, not enough investigating of personal and private matters is done. (.17, —.27)

In general, these individual item analyses confirm, as did the total score correlations, some degree of relationship between F-scale content and ethnocentrism over and above that accounted for by shared acquiescence set. With regard to the components of the F scale, the sex items seem to do very well. The authoritarian and conventionality content is also represented among the successful items. The anti-intraceptive items, although representing one of the most fascinating and plausible of the original syndromes, seem, under these conditions, worthless as predictors of E content. The reason for this is not immediately apparent. This finding should be cross-validated with new populations and item samples before extensive interpretations are made.

SUMMARY

To investigate the effect of acquiescence response set upon the correlations of the E and F scales of *The Authoritarian Personality,* negatively worded sets of items were prepared for each and administered with the original positive items. Evidence of acquiescence bias was sought through intratest analyses and especially through analysis of relationships between the tests. The E scale is the more internally consistent, and by internal criteria, the least susceptible to acquiescence bias. The negative correlation of ethnocentrism and intelligence remains when reversed items are used. While the F scale shows stronger evidence of acquiescence bias, and while the correlation with intelligence seems specific to the original form of the test, the general validity of the syndrome of personality correlates of ethnocentrism is confirmed. However, the component of anti-intraception fails to be confirmed when checked against a set-free measure of ethnocentrism.

Manuscript received: July 17, 1958
Revised manuscript received: February 16, 1959

Loren J. Chapman
Department of Psychology
University of Chicago
Chicago 37, Illinois

REFERENCES

1. Adorno, T. W., E. Frenkel-Brunswick, D. J. Levinson, and R. N. Sanford, *The Authoritarian Personality*, New York: Harper, 1950.
2. Bass, B. M., "Authoritarianism or Acquiescence," *Journal of Abnormal and Social Psychology*, 1955, 51, 616–623.
3. Carey, G. L., A. A. Rogow, and C. Farrell, "The Relationship between the F Scale and Aphorism Usage and Agreement," *Journal of Psychology*, 1957, 43, 163–167.
4. Chapman, L. J., and D. T. Campbell, "Response Set in the F Scale," *Journal of Abnormal and Social Psychology*, 1957, 54, 129–132.
5. Chapman, L. J., and R. D. Bock, "Components of Variance Due to Acquiescence and Content in the F Scale Measure of Authoritarianism," *Psychological Bulletin*, 1958, 55, 328–333.
6. Christie, R., "Authoritarianism Re-examined," in R. Christie and M. Jahoda, eds., *Studies in the Scope and Method of the Authoritarian Personality*, Glencoe, Ill.: The Free Press, 1954, pp. 123–196.
7. Christie, R., and J. Garcia, "Subcultural Variation in Authoritarian Personality," *Journal of Abnormal and Social Psychology*, 1951, 46, 457–469.
8. Christie, R., J. Havel, and B. Seidenberg, "Is the F Scale Irreversible?" *Journal of Abnormal and Social Psychology*, 1958, 56, 143–159.
9. Cohn, T. S., "The Relation of the F Scale to a Response Set to Answer Positively," *Journal of Social Psychology*, 1956, 44, 129–133.
10. Cronbach, L. J., "Response Sets and Test Validity," *Educational and Psychological Measurement*, 1946, 6, 475–494.
11. Cronbach, L. J., "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 1951, 16, 297–334.
12. Gage, N. L., G. S. Leavitt, and G. C. Stone, "The Psychological Meaning of Acquiescence Set for Authoritarianism," *Journal of Abnormal and Social Psychology*, 1957, 55, 98–103.
13. Jackson, D. N., and S. J. Messick, "A Note on 'Ethnocentrism' and Acquiescent Response Sets," *Journal of Abnormal and Social Psychology*, 1957, 54, 132–134.
14. Jackson, D. N., S. J. Messick, and C. M. Solley, "How 'Rigid' is the 'Authoritarian'?" *Journal of Abnormal and Social Psychology*, 1957, 54, 137–140.
15. Leavitt, H. J., H. Hax, and J. H. Roche, "'Authoritarianism' and Agreement with Things Authoritative," *Journal of Psychology*, 1955, 40, 215–221.
16. Lentz, T. H., "Acquiescence as a Factor in the Measurement of Personality," *Psychological Bulletin*, 1938, 35, 659.
17. Levitt, E. E., "The Water-Jar Einstellung Test as a Measure of Rigidity," *Psychological Bulletin*, 1956, 53, 437–470.
18. Lorge, I., "Gen-like: Halo or Reality," *Psychological Bulletin*, 1937, 34, 545–546.
19. Messick, S., and N. Fredericksen, "Ability, Acquiescence, and 'Authoritarianism,'" *Psychological Reports*, 1958, 4, 687–697.
20. Murphy, G., and R. Likert, *Public Opinion and the Individual*, New York: Harper, 1938.
21. Thorndike, R. L., *Personnel Selection: Test and Measurement Techniques*, New York: Wiley, 1949, pp. 345–351.
22. Walker, Helen M., and J. Lev, *Statistical Inference*, New York: Holt, 1953.

Developmental Trends in Small Groups

HUGH PHILP, *University of Sydney*
DEXTER DUNPHY, *University of Sydney*

Continuing interest in social psychology has centered about the problem of determining whether definite steps or "phases" are observable in group problem solving, and whether these, when discovered, can be refined in some way to make problem solving more effective. A number of authors (10, 11, 13, 14) base their analysis of problem solving in groups on Dewey's five steps of effective thinking (8, p. 72). These analyses tend to be more prescriptive than descriptive of actual processes and, as Bales and Strodtbeck point out (4, p. 485), "the distinction between predicting an empirical order of phases under given conditions and prescribing an ideal order in terms of value judgments has not in all cases been clearly drawn."

While theorizing has been comparatively prolific, there has been until recently little experimentation directed toward discovering whether such phases can be in fact detected and measured in any way. Undoubtedly the most significant research that has been carried out along these lines is that of R. F. Bales and associates in the Laboratory of Social Relations at Harvard. Taking his theoretical origins in Weber and Parsons, Bales sees the face-to-face group as a small social system composed of individuals engaged in a process of interaction and communication in such a way that each gains some immediate psychological impression of the others. Group problem solving involves the evolution of a social system, in which each act is seen in terms of its functional relevance to the growing consensus of the group.

Each group is seen as having two basic tasks to perform: (a) to solve the objective problem to which the group is committed, and (b) to build, strengthen, and regulate the group life. Thus all group activities are directed to task goals or system goals; task goals being goals associated with the task the group exists to undertake, and system goals being those which exist to integrate the group relationships. The emerging social structure of a group may be regarded as a system of solutions to the problem of initial instability of interaction which become institutionalized to reduce the uncertainty and unpredictability of others' actions. The group structure in this view develops as individual action patterns become predictable, and it is only as this emerging stability becomes apparent in a group that the group is free to devote its energies primarily toward problem solving in the task area rather than to developing and stabilizing intermember emotional relationships.

On the basis of such theory as this, Bales has developed a set of observa-

tional categories (2) which implies a sequence of functions necessary for progression from a problem situation to consensus or agreement. Briefly, the sequence is regarded as a system of interaction between members over time which must pass through successive phases of communication, evaluation, control, decision, and tension reduction.

In their experiments Bales and his associates (3, 4, 5, 12) have used carefully contrived groups discussing specially constructed artificial problems. In this way their hope has been to isolate the particular system properties while allowing personality and other variables to operate in a random fashion. As a consequence it has been shown that, under the particular limiting conditions, phase movements do occur (4). It remains an empirical problem to determine whether with other kinds of group—"natural" groups for example—and other kinds of problem—those with high ego-involvement for instance—similar phases occur.

In the research to be reported here we were interested to discover (a) whether contrived groups employed on complex problems directly related to their course work would in fact pass through a similar series of distinguishable phases, and (b) whether the sex of the participants would influence such a phase pattern in any way. In addition we were interested in discovering some of the characteristics of the emergent social structure of the groups. That is, we were concerned with one particular kind of variant on the "pure" situation employed by Bales. While maintaining the "contrived" rather than the "natural" group situation, we decided to examine the emergence of phases under conditions in which solution of a particular problem had ego relevance.

PROCEDURE

The subjects for the research were male and female undergraduates, all day and evening students of the course labeled "Education II" in the Faculty of Arts at the University of Sydney. There were 88 students in all, comprising 38 females and 50 males. The students were assigned, by use of random number tables, to eleven discussion groups, each group consisting of eight members. The sex factor was controlled by a distribution which systematically varied the proportion of males to females over the groups as shown in Table 1.

Eight discussion sessions of 40 minutes each were held over a period of ten weeks on a problem closely related to the course in education. The topic assigned to the groups for discussion read as follows: "What would you regard as necessary elements of a general education in Australian schools at the secondary level? Outline as clearly as possible the criteria you would use in making your choice." Groups were instructed to "discuss the problem and endeavor to reach conclusions to be embodied in a report to be sub-

mitted at the close of the discussion period." When the reports were received at the close of the eight sessions they were given in turn to four lecturers in Education who were instructed:

1. To rank the reports according to defined criteria.
2. To mark them individually as "pass," "credit," or "fail" in relation to their standard as university work in Education II, a third year course.
3. To shuffle the reports before handing them on to the next marker.

The observers for the experiment consisted of members of the departmental staff and honors students. Five training sessions of about 1½ hours each

TABLE 1

Distribution of Males and Females over Groups

									Repeat Proportions		
Group No.	1	2	3	4*	5	6	7	8	9	10	11
No. of males	0	1	2	3	5	6	7	8	5	6	7
No. of females	8	7	6	5	3	2	1	0	3	2	1
Total:	8	8	8	8	8	8	8	8	8	8	8

* Note: Complete records were not available for group 4. It is therefore omitted from subsequent tables.

were held prior to the experimental period. In addition a retraining session was held during the experimental period. Observer reliability was checked by having all observers score simultaneously a segment of interaction at a practice session. The scores were then tabulated and compared using the method of allowances. It was found that, at the 5 per cent level of confidence, there were no significant differences in the numbers of acts scored in each category by the observers and so they may be held to be scoring reliably.

The method of observation used throughout the experiment was, with some modifications to be described below, the procedure outlined by Bales in *Interaction Process Analysis* (1) and elsewhere. The categories he has evolved for the study of small groups represent a consistent, complete, and systematic set of carefully defined concepts which can be used for observing any kind of face-to-face group. The scheme has the advantage of giving numerical scores that can be analyzed by statistical methods and this facilitates the comparison of large amounts of data. The method of scoring is continuous and inclusive, the observer recording the interactive processes continuously, scoring every perceivable act as it occurs according to its context and significance for the social system. Each unit of observed behavior is classified into one of a set of 12 categories. This set of categories includes behavior of four types: questions, answers, negative reactions, and positive reactions.

As each unit of communication is expressed, the observer records its appropriate number from the list, and indicates the originator of the act and the person to whom the act is directed. The classification of acts is essentially a "surface" analysis in that the observer scores in terms of the act or two preceding and, perhaps, following the act under observation. In general, the observer tries to adopt the position of the alters to whom the act is directed in order to determine what the instrumental or affective significance of the act is to the group members.

There were two main differences in observational procedure between Sydney and Harvard: (a) In Sydney we had no soundproof room and group members were aware of the physical presence of the observers. However, the observers operated as mechanically as possible and generally groups appeared quite uninhibited by them. At Harvard groups were acquainted with the presence of the unseen observers and so were also conscious of being observed. (b) In the Sydney research no mechanical recording apparatus was used, acts being allotted to a particular phase and not recorded in strict sequence as this was irrelevant to the research. However, the method of scoring used was not difficult and the observers became quite proficient during training sessions.

With these minor differences, the method used was a replication of that used by Bales and his associates at Harvard and so comparison between the Sydney and Harvard groups is possible. One major point of difference in terminology must be noted here however. With the research on phase movements conducted by Bales and Strodtbeck (4), observation was confined to groups discussing problems which were solved in a single session only, but the problem allotted to the Sydney groups occupied eight sessions. Thus there is a considerable difference in the comparative length of the phase unit to the whole problem-solving period—Bales divided a single session into three and we into four, but for Bales's groups a phase was one third of the total problem-solving process without replication, while for our groups it was one quarter, with each phase replicated eight times.

RESULTS

For the purposes of the present analysis, the 12 categories were not treated separately, but were divided into their four main areas: categories 1–3 (positive social-emotional reactions), 4–6 (attempted answers), 7–9 (questions), and 10–12 (negative social-emotional) represented on the tables as A, B, C, and D respectively. Data were then recorded in terms of the number of acts in each of these areas for each phase of each session.

The number of acts made by a group in any one session or phase of a session may be regarded as the "score" of the group for that session or phase.

The total N is thus not the number of acts, but the number of classifications of acts into sets of scores—i.e., we had, finally, 1280 "scores" each of which represented the number of acts made by a group in a particular category during a specific phase of a session.

The complete table was then analyzed by means of analysis of variance. The residual variance was used as an error term for testing of second-order

TABLE 2

Analysis of Variance—Summarized Results

Source of Variation	Degrees of Freedom	Mean Square	F	p
Groups	9	1,077	2.20 N.S.	
Categories	3	57,696	118	<.001
Sessions	7	1,324	3.195	<.01
Phases	3	630		
GPS x Cats.	27	489	1.77	<.05
GPS x Sess.	63	207		
GPS x Phas.	27	91		
Cats & Ses.	21	337		
Cats & Ph.	9	117		
Sess. & Ph.	21	659	1.59 N.S.	
GPS x C x S	189	71		
GPS x C x P	81	299	1.08 N.S.	
Cats x S x P	63	414	1.50	<.01
GPS x S x P	189	58		
Residual	567	277		
Total	1279	—		

interactions. In the event of a significant second-order interaction, the mean square for this was used as the error term for testing of appropriate first-order interactions. Main effects were tested against the mean square of appropriate significant first-order interaction, e.g., the main effect "categories" was tested against the significant first-order interaction groups-categories.

The results of the analysis of variance show significant differences attributable to categories and sessions of the main effects, and to groups-categories, and categories-sessions-phases in interaction effects. Other variables and interaction effects are not responsible for significant differences in the data.

Sessions

The analysis shows quite clearly that there is a significant change in the number of acts from session to session. Table 3 gives the total number of acts per session.

TABLE 3

Number of Acts per Session

Session	1	2	3	4	5	6	7	8
No. of acts	1941	1692	2417	3004	2947	2653	2692	2460

Table 3 shows a drop in the number of acts in session 2, a sharp rise to session 4, followed by a tendency to fall with succeeding sessions. It seems then that activity was lowest in session 2, highest in session 4, and tended to fall steadily after reaching this point. However, activity in session 8 was still well above that in session 1.

Categories

The numbers of acts falling into the four different categories were also shown to be significantly different, i.e., more acts fell into some categories than into others. Thus of the 19,806 acts recorded over all sessions and for all groups, each category received numbers as shown in Table 4.

By far the greatest proportion of acts (50 per cent) fell into the attempted answers sector (category B); positive social-emotional acts (category A) received the next highest total, closely followed by the questions area (category C). Negative social-emotional unity-destroying behavior accounted for least of all the acts. Category B contained about 56 per cent of the acts, category A about 20 per cent, category C about 17 per cent and category D only about 6½ per cent of the total number of acts over all groups. There were about three times as many attempted answers as questions, and about three times as many positive reactions as negative reactions.

Groups-categories

The analysis also reveals that there are significant differences in the numbers of acts falling into each category from group to group; i.e., that different groups tend to use different categories to a significantly different

TABLE 4

Number of Acts per Category

Category A	3,964
Category B	11,164
Category C	3,380
Category D	1,298

extent. Table 5 shows the percentage of acts falling into each of the four categories for each group.

These figures reveal that in all groups about 50 per cent or more of acts fall into category B, the attempted answer sector of the task area, while category C receives only from about 11 per cent to about 24 per cent of the total number of acts. Similarly, in the social-emotional area there is in all groups a higher percentage of positive reactions than negative reactions, of group-oriented activity as against self-oriented activity. However, there is more

TABLE 5

Groups and Categories

Categories		A	B	C	D
Groups:	1	25.7	49.1	21.3	3.9
	2	10.5	56.8	23.8	8.9
	3	18.4	65.7	13.0	2.9
	5	21.4	60.8	10.9	6.8
	6	21.4	54.6	17.4	6.5
	7	20.6	53.6	19.5	6.3
	8	19.1	56.6	17.6	6.6
	9	16.8	56.2	14.5	12.5
	10	20.3	57.2	18.0	4.6
	11	25.1	52.1	16.5	6.2
Mean:		19.93	56.27	17.25	6.52

variation from group to group in the extent to which this is so than in the task area. For instance, in group 2 the percentage of acts falling into category A hardly exceeds the percentage falling into category D (10.5 to 8.9) while in group 1 the percentage of acts in category A is six times as great as the percentage in category D. This means that in all groups overt showing of solidarity exceeded unity destroying behavior, but that there is a large amount of variation from group to group in the extent to which this is so.

Categories-phases-sessions

The analysis shows this interaction to be significant, so significant in fact as to mask any phase movement in categories within any particular session. There is, that is to say, a highly significant tendency in all groups to use different categories in different phases as the number of sessions increases. It matters, in terms of the relative emphasis on the categories, whether a phase is phase 1 in session 1 or phase 1 in session 6. Examination of the basic data suggests that in early sessions a greater proportion of positive social-

emotional type acts (category A) occur in phases 1 and 2, than in phases 3 and 4, whereas in later sessions, social-emotional type acts tend to occur in phases 1 and 4.

A significant phase movement thus exists, but the relative emphasis on the various categories varies according to the phase within the *total* process of problem solving. Since in our data this extended over eight sessions the detailed pattern is a little different from that of Bales, but the *general* pattern is very similar.

It is not possible to determine from this study whether this phase movement is predictable or not, for there is only one case (i.e., the whole process involving all groups). But within this one case a phase movement does manifest itself; furthermore, it can be seen that the secondary hypothesis which stated that the phase movement is not affected by the sex composition of the groups is verified. There were no significant differences between groups, in fact the only place where the group factor is significant is with groups-categories, and the variations here bear no apparent relationship to the sex-composition of the groups. Thus groups, and so sex (for sex was varied from group to group), did not affect the phase movement.

Emotionality of Group Process and Quality of Group Production

According to the theory briefly outlined at the beginning of the article, successful group problem solving depends upon reaching a satisfactory solution to the problem of interpersonal relationships among the group members. This is a group's primary concern on being thrown together for the first time and takes precedence over the objective problem being dealt with. Once stable relationships have been achieved the group is free to concentrate its attention on the task in hand, in this theory. One would expect then that groups which show a high percentage of acts in the social-emotional area of group life (i.e., groups which have not satisfactorily solved the problems of inter-personal adjustments) would be far less effective in solving the objective problem. It will be remembered that each group was required to furnish a report on the group's decisions on the discussion problem at the end of the eight sessions, and that these reports were ranked by four university lecturers. The rankings on content of report for each group are given in Table 6 below. By adding category A to category D we can determine which groups expended the greatest percentage of acts on the emotional aspects of group process. The rankings on social-emotionality are also given in Table 6. We are now in a position to compare the two rankings and may hypothesize that groups which rank high on social-emotionality will rank low on their final reports.

TABLE 6

Social-Emotional and Content Rank Compared

	Social-Emotional Rank	Content Rank	
Rank 1	Group 11	Group 8	
2	" 1	" 3	
3	" 9	" 7 & 10	High
4	" 5		
5	" 6	" 2	
6	" 7	" 11	
7	" 8	" 1	
8	" 10	" 5	Low
9	" 3	" 9	
10	" 2	" 6	

It will be seen from this that those groups which are ranked high on a social-emotional criterion tend to be those which are ranked low on quality of final report (Spearman's $\rho = -.46$). A later article will discuss the implications of this finding.

DEVELOPMENTAL TRENDS IN GROUP STRUCTURE

A contrast of the proportion of acts devoted to each of the four categories session by session shows up some interesting changes which occurred over time in the developing structure of the groups. The combined scores for all groups converted into percentages for each session are shown in Table 7.

The scores were tested for trend using the method of allowances but it was found that there were no significantly consistent trends. Each percentage was then compared with the one next to it to determine whether or not they differed significantly. The significances are shown on Table 7 [$>$ means significantly greater at the 5% level].

TABLE 7

Interaction Trends Over Time for the Sydney Groups Combined

Categories	Percentage Rates by Session							
	1	2	3	4	5	6	7	8
A	20.2 $>$	14.4 $<$	17.9 $<$	22.7	20.1	21.4	20.3	20.7
B	53.1 $<$	56.8	56.9 $>$	54.2 $<$	57.0	56.7	58.1	57.8
C	18.3 $<$	20.1 $>$	18.2	16.5	15.8	16.6	16.0	16.7
D	8.4	8.7	7.0	6.6	7.1 $>$	5.3	5.6	4.8
A-D	11.8 $>$	5.7 $<$	10.9 $<$	16.1 n.s.	13.0 n.s.	16.1 n.s.	14.7 n.s.	15.9

This reveals some significant differences between the percentages of acts in the categories A, B, and C in the early sessions. It indicates that significant variations occur in the first few sessions, but that from session 5 onward with one minor exception variations do not occur to a significant extent in any category. It will be remembered that, according to the theory outlined earlier, a stable structure tends to develop over time in the social system of the problem-solving group. These results are added support for the theory for they show that large differences in the extent to which different categories are used in the early sessions give way to a developing stability with time. It should be noted that this stability tends to maintain the task-oriented activity at a high level.

Social-emotional Trends

In category A (positive reactions) there is a significant decrease in the percentage of acts in session 2, and a significant rise again to session 4. After this point there is a return to approximately the same level as in session 1 and a tendency to maintain this level. The major instability in this category clearly occurs in session 2 where there is a marked decrease in positively oriented social-emotional behavior. In category D (negative reactions) there is a slight, but not a significant rise, in session 2 followed by a slow drop by nonsignificant steps (except between sessions 5 and 6). There is, however, a significant difference between the level of session 1 and that of session 8 which means that over the eight sessions there was a significant decrease in overt negative social-emotional behavior that indicates a growing consensus in groups over time and an increasing acceptance of group norms of behavior.

Task Trends

Category B (attempted answers) shows a significant rise in the percentage of acts from session 1 to 2, this level was maintained in session 3, decreased significantly in session 4, increased again in session 5, and tended to maintain a steady level from then on. The drop in task-oriented acts in session 4 can probably be accounted for by the completion at the beginning of this session of the interim reports. This was apparently marked by a drop in task-oriented activity and a rise in expressions of solidarity and satisfaction. However, with the prospect of producing a final report within a few sessions the upward trend in task-oriented activity is resumed in session 6.

Category C (questions) shows a significant rise in session 2 followed by a drop in session 3 and a tendency from that point to maintain stability.

It has been shown that it is in session 2 that the greatest changes in relative emphasis on categories took place. This session was marked by a higher

proportion of questions, a high proportion of attempted answers and negative reactions, and a major decrease in positive, unity-producing behavior.

Comparison with Research in the United States

In an article "Developmental Trends in the Structure of Small Groups" (12), Heinicke and Bales give details of groups studied at Northwestern and Harvard Universities and scored using Interaction Process Analysis. The groups consisted of male undergraduates discussing human relations problems. The groups had to reach conclusions and make recommendations. Sessions were of 40 minutes duration held at weekly intervals; four sessions were held at Harvard, six at Northwestern. In Table 8 the authors present figures for the combined groups for the first four sessions.

TABLE 8

Interaction Trends Over Time for Northwestern and Harvard Groups Combined

Categories	Percentage Rated by Session				Level of Significance
	1	2	3	4	
A	26.9	27.1	27.8	33.9	**
B	54.6	49.7	51.3	43.4	
C	4.3	3.9	3.9	3.7	
D	10.5	16.0	11.8	12.0	**
A-D	15.0	6.8	14.5	19.8	**

NOTE: Two asterisks (**) stand for the .01 level of significance.

The figures for the United States groups reveal a somewhat different pattern from the Sydney groups though they were working under roughly comparable conditions. The figures for the first four sessions show decreasing rates in category B compared with increasing rates in the Sydney groups. Bales's groups show much lower percentages in category C.

The social-emotional categories show similar differences. The United States groups do not show the sharp drop in positive emotional reactions in the second session shown by the Sydney groups, although they do show a similar rise in the third and fourth sessions. Where the United States groups show a sharp rise in negative emotional reactions in session 2 the Sydney groups register only a slight nonsignificant rise.

However, the subtraction of category D from category A reveals a remarkable similarity between the Sydney and Harvard/Northwestern groups.

Table 9 shows clearly that for both sets of groups there was a significant

TABLE 9

A-D

Session	1	2	3	4
U.S. groups	15.0	6.8	14.5	19.8
Sydney groups	18.8	5.7	10.9	16.1

decrease in "excess" positive social-emotional reactions in the second session, that in the third session the former level was regained, and that the fourth session showed a significant increase above this again. Further figures for the Sydney groups reveal that this level tends to be maintained in subsequent sessions. It becomes apparent from this analysis that, in problem-solving groups such as these, session 2 is clearly the session of the greatest conflict, and that from this point onward there is an increasing proportion of positive social-emotional acts as against negative social-emotional reactions.

CONCLUSION

It would seem then that the general theoretical position of Bales is supported by the Sydney research. Our results, *mutatis mutandis*, are of the same order except for the relative weights of categories B and C in the early sessions. This may reasonably be accounted for in terms of the relative difficulty of the problems, for patently where the issue is more complex more elucidatory questions have to be asked. Then, too, in the Harvard problems almost all the data necessary for solution was common to all participants, whereas in the Sydney set, this was not so; the members of each group had widely varying academic and professional backgrounds.

The social-emotional differences in the second session are of somewhat greater interest. The combined Sydney groups' scores showed a drop in positive social-emotional acts, but no corresponding rise in negative social-emotional acts, while the United States groups showed no positive change but an increase in the negative social-emotional area. That is, the Sydney groups, in the second session, were much more concentrated in the task area than their American counterparts who were still concerned with solving their problems of getting along together. At least two explanations suggest themselves: one is that there is some "national" difference between the groups, and the other, which seems more likely, is that with greater ego-involvement in the task itself, the necessity for getting down to it tends to override social-emotional issues. Yet this is only a *relative* matter. For the Sydney groups as for the American ones, the second session is the one in which the most

conflict occurs. Beyond this a new phase of cooperation and concentration in the task area occurs.

Manuscript received: November 13, 1956
Revised manuscript received: August 4, 1958
Hugh Philp
Department of Education
University of Sydney
Sydney, N.S.W., Australia

REFERENCES

1. Bales, R. F., *Interaction Process Analysis,* Cambridge, Mass.: Addison-Wesley, 1950.
2. Bales, R. F., "A Set of Categories for the Analysis of Small Group Interaction," *American Sociological Review,* 1950, 15, 257–263.
3. Bales, R. F., and H. Gerbrands, "The Interaction Recorder; An Apparatus and Check List for Sequential Content Analysis of Social Interaction," *Human Relations,* 1948, 1, 456–463.
4. Bales, R. F., and F. L. Strodtbeck, "Phases in Group Problem Solving," *Journal of Abnormal and Social Psychology,* 1951, 46, 485–495.
5. Borgatta, E. F., and R. F. Bales, "The Consistency of Subject Behavior and the Reliability of Scoring in Interaction Process Analysis," *American Sociological Review,* 1953, 18, 566–569.
6. Borgatta, E. F., and R. F. Bales, "Task and Accumulation of Experience as Factors in the Interaction of Small Groups," *Sociometry,* 1953, 26, 239–252.
7. Cartwright, D., and A. Zander, *Group Dynamics—Research and Theory,* Row, Peterson, Evanston, Ill., 1953.
8. Dewey, J., *How We Think,* Boston, Mass.: Heath, 1933.
9. Dixon, W., and F. Massey, *Introduction to Statistical Analysis,* New York: McGraw-Hill, 1951.
10. Elliot, H. S., *The Process of Group Thinking,* New York: Association Press, 1938.
11. Hall, D. M., *The Dynamics of Group Discussion,* Illinois: Interstate Printers and Publishers, 1950.
12. Heinicke, C., and R. F. Bales, "Developmental Trends in the Structure of Small Groups," *Sociometry,* 1953, 16.
13. Lasker, B., *Democracy through Discussion,* New York: Wilson, 1949.
14. McBurney, J. H., and K. G. Hance, *Discussion in Human Affairs,* New York: Harper, 1950.
15. Slater, P. E., "Role Differentiation in Small Groups" in A. P. Hare, E. F. Borgatta, and R. F. Bales, *Small Groups,* New York: Knopf, 1955.
16. Snedecor, G. W., *Statistical Methods,* Ames, Iowa: Iowa State College Press, 1950.

Negro and White Reactions to Racial Humor [1]

RUSSELL MIDDLETON, *Florida State University*

A number of studies of humor have been conducted during the last three decades which have attempted to isolate "sense of humor" as a general personality trait and to relate various other social and psychological traits to "sense of humor." (See, for example, 2, 8, 10, 14, and 19.) Some of the more sophisticated studies have recognized that humor is a function of the specific cultural setting and that there are great differences in cultural definitions of humor (4, 6, 18). Thus, an anecdote which is highly amusing to a person with one cultural background may be meaningless or even repugnant to a person who has internalized different cultural values. There is more than a little truth in the maxim of the field anthropologist that when the investigator has progressed to the point of understanding the jokes of the native people as much as half of the time, it is possible for him to leave the field confident in the knowledge that he has a thorough understanding of the basic elements of the culture.

Relatively little research has been devoted to racial or social class subcultural variations in humor. Neither has much attention been given to the socio-psychological study of jokes of ridicule which are directed against a minority group or an out-group, beyond the simple recognition that such jokes function as a conflict mechanism in intergroup relations (cf. 12). In one of the rare investigations of subcultural differences in the evaluation of jokes disparaging a minority group, Wolff, Smith, and Murray compared the "mirth responses" of approximately six Jewish subjects and nine Gentile subjects to eight jokes disparaging Jews and to eight other jokes introduced as a control. Although both Jews and Gentiles rated the anti-Semitic jokes as less funny than the control jokes, the anti-Semitic jokes were rated considerably higher by the Gentile subjects than by the Jewish subjects. The authors argued that a positive response to a race-disparagement joke is an indication of negative sentiments toward the disparaged object. They concluded then that ". . . individuals . . . *do not* laugh at the misadventurous experiences or discreditable aspects of objects for which it is certain they possess positive sentiments and . . . conversely, they *do* laugh at the expense of unaffiliated objects. . . ." (20, pp. 361–364).

In spite of the surface plausibility of the conclusions of Wolff, Smith, and

[1] I am indebted to John Moland of Florida A & M University for assistance in the collection of the data for this study and to the Center for Social Research of Florida State University for aid in processing the data.

Murray, paradoxically there is a great deal of evidence that Jews themselves quite frequently tell self-disparaging jokes which are based upon the traditional stereotype of the Jew as shrewd, mercenary, grasping, clannish, and aggressive (5, pp. 704–707; 11, pp. 168–170; 15, p. 75). In a recent study of jokes told by white and Negro university students, it was also found that the Negroes told more "anti-Negro" jokes based upon the traditional stereotype of the Negro than did the whites. Furthermore, the Negro students told almost four times as many anti-Negro jokes as "anti-white" jokes—that is, racial jokes at the expense of the whites (12).

This paper reports the results of a study undertaken to compare the appeal of racial jokes—both anti-Negro and anti-white—to matched groups of Negroes and whites. A further attempt was made to determine some of the variables related to favorable or unfavorable reactions to the racial jokes.

On the basis of the Wolff-Smith-Murray theory, the following hypotheses were formulated:

1. Whites react more favorably to anti-Negro jokes than do Negroes; Negroes react more favorably to anti-white jokes than do whites.

2. Authoritarianism is positively related to favorable responses to anti-Negro jokes among whites and to anti-white jokes among Negroes (cf. 13).

3. Authoritarianism is negatively related to favorable responses to anti-white jokes among whites and to anti-Negro jokes among Negroes.

4. Acceptance of elements of the stereotype of the Negro is positively related to favorable reactions to anti-Negro jokes based upon those stereotypes.

An exploratory null hypothesis regarding social class differences was also formulated: there is no significant difference between social classes in reactions to the jokes of the various types for each of the racial groups.

METHOD

The matched pairs technique was utilized to secure a Negro experimental group and a corresponding white experimental group. Subjects were recruited from among students at two state universities—one for whites and one for Negroes. Each Negro student was matched to a white individual on the following personal characteristics: age, sex, education, social class (parental occupation), and predominant residence within or outside the South. Fifty pairs or a total of 100 subjects were secured. To permit comparisons within each of the racial groups between middle and lower class subjects, matching also was carried out between class groups within each race. Thus there were 25 persons in each of the race and class groupings who were matched on other personal characteristics.

Eighteen jokes of four different types were selected for presentation to the subjects: six anti-Negro jokes, six anti-white jokes, three punning jokes, and

three pointless jokes. All the jokes were taken from published sources which were not likely to be familiar to the subjects. The pointless jokes were, however, altered from their original form by removing the "punch lines" and substituting innocuous and pointless lines. The puns and pointless jokes were introduced as a further control measure, since one would not expect very great race differences in reactions to these types of humor. In order to accept the hypotheses then it is necessary that when significant differences in reactions to the anti-Negro and anti-white jokes are present, there should be no significant difference between groups for the control jokes.

The 18 jokes were presented to the subjects as a paper-and-pencil test in a classroom situation. They were asked to rate each of the jokes for degree of funniness according to a 7-point rating scale ranging from "extremely funny" to "not funny." For summary tabular presentation the ratings are dichotomized into "relatively funny," including the ratings of average funniness and above, and "relatively not funny," including the ratings of below average funniness. Although the paper-and-pencil method of presenting the jokes and securing ratings does present a highly artificial social situation, the experience of most other investigators has been that variations in the method of presentation and in the method of measuring responses have had relatively little effect upon the results. The paper-and-pencil test then was selected for use both because of its simplicity and because of its exclusion of many relatively uncontrolled variables which were found in other methods.

Immediately after the subjects completed the ratings of the 18 jokes, Adorno's F-scale for authoritarianism was administered (1). Although the F-scale has come under attack for a number of methodological defects, Hyman and Sheatsley point out that it is acceptable to use the scale as a disguised instrument for the measurement of prejudice or ethnocentrism (7, p. 75). For some purposes of analysis the subjects were divided into a group of high authoritarianism with an item mean score of 4.25 or above and a group of low authoritarianism with an item mean score of less than 4.25.

Adjoined to the F-scale questions was a series of four questions designed to measure the relative acceptance or rejection of four elements of the Negro stereotype: love for certain foods such as chicken, immorality in sexual behavior, dishonesty, and laziness.

A composite rating score for each category of joke was calculated for each individual by summing the numerical ratings of the jokes. On the basis of these scores, differences between groups in reactions to humor were subjected to the Mann-Whitney *U* test to determine the probability that the groups could have been drawn from the same population. The Mann-Whitney *U* test is nonparametric and assumes only ordinal measurement (16, p. 116). Spearman rank correlations were utilized for measuring the relation between authori-

tarianism and reactions to the various types of humor, and the *t* test of significance was computed for the rank correlations. The probability level used for rejecting the null forms of the hypotheses was .05.

RESULTS

A gross comparison of the ratings made by white and Negro subjects for each of the types of jokes is presented in Table 1. There was no significant difference between whites and Negroes in reactions to anti-Negro jokes or to the control jokes, either the puns or pointless jokes. Thus the data suggest

TABLE 1

Per Cent of Ratings "Relatively Funny," by Race of Respondents and Type of Joke

	White		Negro	
Type of Joke	N	Per Cent	N	Per Cent
Anti-Negro	300	54.3	300	53.7
Anti-white	300	40.0	300	51.3
Puns	150	37.3	149	29.5
Pointless	149	25.5	149	30.2

that Negroes found the anti-Negro jokes quite as funny as did the whites. With regard to the anti-white jokes, however, the Negroes reacted more favorably than the whites. These comparisons must be accepted with caution, however, for the important variable of authoritarianism or ethnocentrism is left uncontrolled, and the Negro subjects in general represent a considerably higher level of authoritarianism. The mean of the item mean scores on the F-scale for Negroes was 4.56 and for whites was 3.65.[2]

In Table 2 a control for authoritarianism is introduced in the comparison of the reactions of whites and Negroes to each type of joke. Although among the highly authoritarian persons the whites reacted more favorably to anti-Negro jokes than did the Negroes, just as was hypothesized, among those with relatively low scores on the F-scale, the Negroes reacted more favorably than the whites. This latter deviation from expectations is probably to be explained in terms of feelings of guilt and a general "touchiness" with regard to race relations on the part of the nonauthoritarian whites.

In the case of persons with relatively low scores on authoritarianism, Negroes responded more favorably to the anti-white jokes than whites, as was

[2] Smith and Prothro found a mean of 4.68 for Negroes and 3.86 for whites on the F-scale in a study of subjects taken from the same two universities utilized in this study. The subjects were, however, neither randomly selected nor matched to eliminate the influence of sex and social class variables (17).

TABLE 2

Per Cent of Ratings "Relatively Funny," by Race and Authoritarianism of Respondents and Type of Joke

	Low Authoritarianism				High Authoritarianism			
	White		Negro		White		Negro	
Type of Joke	N	%	N	%	N	%	N	%
Anti-Negro	222	46.4	90	63.3	78	76.9	210	49.5
Anti-white	222	34.7	90	55.6	78	55.1	210	49.5
Puns	111	35.1	44	36.4	39	43.6	105	26.7
Pointless	110	20.9	44	25.0	39	38.5	105	32.4

expected. In the highly authoritarian group, however, there was no significant difference between the responses of the whites and the Negroes to the anti-white jokes. The failure of highly ethnocentric whites to react in a markedly negative fashion to jokes which ridicule themselves suggests that there was a tendency for the subjects to identify and equate both anti-Negro jokes and anti-white jokes as "racial jokes." Individuals, it would seem, tend toward more or less uniform reactions to racial jokes regardless of the nature of the content.

There was no significant difference between whites and Negroes in either the high or the low authoritarian groups in reactions to the control jokes.

In Table 3 are presented the Spearman rank correlations between authoritarianism and positive responses to each type of joke. There was a correlation of +.436 between authoritarianism and appreciation for anti-Negro jokes among the whites, but there was an almost identical correlation of +.437 for anti-white jokes among the whites. The correlation between favorable reactions to anti-Negro jokes and favorable reactions to anti-white jokes was +.499. Once again the evidence indicates the identification of anti-Negro and anti-white jokes as racial jokes, to which there are more or less uniform reactions. Correlations for the control jokes among the whites and for all types of jokes among the Negroes were low, and none of these was significant.

TABLE 3

Spearman Rank Correlation of Authoritarianism and Positive Responses to Jokes, by Race of Respondents and Type of Joke

Type of Joke	White	Negro
Anti-Negro	+.436*	−.086
Anti-white	+.437*	+.071
Puns	+.008	−.084
Pointless	−.013	−.021

* Significant at .05 level.

The hypothesis that persons tend to react more favorably to an anti-Negro joke when they accept the validity of the stereotype upon which the joke is based may be examined with reference to the data presented in Table 4. Two jokes based upon each of four different aspects of the stereotype of the Negro were included in the study. Those who accepted a particular stereotype and those who rejected the stereotype may be compared in their reactions to the two jokes based upon the stereotype. Although statistical tests of significance cannot properly be applied to the data, since other variables are not controlled,

TABLE 4

Per Cent of Ratings of Anti-Negro Jokes "Relatively Funny," by Race of Respondents and Acceptance of Aspects of Negro Stereotype

	White				Negro			
	Accepters		Rejecters		Accepters		Rejecters	
Stereotype	N	%	N	%	N	%	N	%
Love of chicken	60	43.3	40	35.0	60	48.3	40	45.0
Sexual laxity	58	79.3	42	78.6	32	71.9	68	63.2
Dishonesty	50	44.0	50	36.0	22	40.9	78	48.7
Laziness	64	48.4	36	36.1	24	45.8	76	48.7

an inspection of the table indicates that there are no great or consistent differences in the reactions of the "accepters" and the "rejecters" among either whites or Negroes. There is apparently operative what Coleridge spoke of as ". . . that willing suspension of disbelief for the moment, which constitutes poetic faith." Thus, even if a person does not accept the validity of a stereotype, he may be willing to suspend his disbelief temporarily in order to enjoy the humor of the joke. This conclusion is highly tentative, however, and requires testing with more factors controlled and a larger number of cases.

Social class differences within each racial group in reactions to humor are shown in Table 5. Individuals were classified as middle class or lower class on the basis of whether their parents were in white collar or blue collar occupations. It should be recognized, however, that among university students even those with lower class backgrounds may take the middle class as a reference group. Consequently these data should be regarded as exploratory.

Among the whites there were no significant differences between the responses of middle class and lower class subjects to any of the types of jokes. Middle class Negroes, on the other hand, responded more favorably than lower class Negroes to both the anti-Negro and the anti-white jokes. Their more favorable response to the anti-white jokes may possibly be explained in terms of their greater feelings of relative deprivation at the hands of the whites. The denial

TABLE 5

Per Cent of Ratings "Relatively Funny," by Race and Social Class of Respondents and Type of Joke

	White				Negro			
	Middle Class		Lower Class		Middle Class		Lower Class	
Type of Joke	N	%	N	%	N	%	N	%
Anti-Negro	150	59.3	150	49.3	150	62.0	150	45.3
Anti-white	150	46.7	150	33.3	150	60.0	150	42.7
Puns	75	36.0	75	38.7	74	35.1	75	24.0
Pointless	74	27.0	75	24.0	74	31.1	75	29.3

of privileges ordinarily accruing to middle class status makes the middle class Negroes more receptive to jokes ridiculing the whites responsible for the denial. Partly because of the refusal of whites to recognize class distinctions within the Negro group, middle class Negroes also tend to be highly class conscious and seek to mark themselves off sharply from lower class Negroes (3, pp. 415–419). The middle class Negroes may thus interpret the ridicule in anti-Negro jokes as directed not at Negroes in general but at the behavior of lower class Negroes. The more favorable reaction of the middle class Negroes to both anti-white and anti-Negro jokes may also be related to the lesser fear of retaliation for responding to the recognized aggression of the jokes (cf. 9).

CONCLUSION

Some of the hypotheses based upon the theory of Wolff, Smith, and Murray that individuals tend to laugh at the expense of unaffiliated objects and not at the expense of objects for which they possess positive sentiments are confirmed by this study, but as many more must be qualified or rejected. In gross comparisons Negroes did react more favorably than whites to anti-white jokes, but the Negroes apparently found the anti-Negro jokes just as funny as the whites. The earlier finding that Negro students tell more anti-Negro jokes than do white students is not so surprising in light of this evidence (12). When authoritarianism is controlled in racial comparisons, additional qualifications must be introduced in the hypotheses. In the low authoritarian group the Negroes reacted more favorably to anti-Negro jokes than the whites, probably due to feelings of guilt by the nonauthoritarian whites. In the highly authoritarian group there was no significant difference between Negroes and whites in their reactions to the anti-white jokes, although it had been hypothesized that the Negroes would react more favorably.

Correlations between authoritarianism and favorable reactions to jokes were low and insignificant except in the cases of the anti-Negro jokes and the anti-

white jokes among the white subjects. There was a moderately high positive association for both types of jokes, indicating a tendency for whites to equate anti-Negro and anti-white jokes as "racial jokes," to which they react in a uniform pattern.

There was little or no evidence to support the hypothesis that persons who accept the validity of the Negro stereotype react more favorably to anti-Negro jokes based upon that stereotype than do persons who do not accept the stereotype.

Finally, social class differences in reactions to humor were significant only among the Negroes, with middle class persons reacting more favorably than lower class persons to both types of racial jokes.

Manuscript received: November 24, 1958
Revised manuscript received: February 18, 1959
Russell Middleton
Department of Sociology
Florida State University
Tallahassee, Florida

REFERENCES

1. Adorno, T. W., *et al.*, *The Authoritarian Personality*, New York: Harper, 1950.
2. Cattell, R. B., and L. B. Luborsky, "Personality Factors in Response to Humor," *Journal of Abnormal and Social Psychology*, 1947, 42, 402–421.
3. Davie, M. R., *Negroes in American Society*, New York: McGraw-Hill, 1949.
4. Eysenck, H. J., "National Differences in 'Sense of Humor': Three Experimental and Statistical Studies," *Character and Personality*, 1944, 13, 37–54.
5. Freud, Sigmund, "Wit and Its Relation to the Unconscious," in A. A. Brill, ed., *The Basic Writings of Sigmund Freud*, New York: Random House, 1938.
6. Heim, A., "An Experiment on Humour," *British Journal of Psychology*, 1936, 27, 148–161.
7. Hyman, H. H., and P. B. Sheatsley, "'The Authoritarian Personality'—A Methodological Critique," in Richard Christie and Marie Jahoda, eds., *Studies in the Scope and Method of "The Authoritarian Personality,"* Glencoe, Ill.: The Free Press, 1954.
8. Kambouropoulou, P., "Individual Differences in the Sense of Humor," *American Journal of Psychology*, 1926, 37, 268–278.
9. Karon, B. P., *The Negro Personality*, New York: Springer Publishing Co., 1959.
10. Landis, C., and J. W. H. Ross, "Humor and Its Relations to Other Personality Traits," *Journal of Social Psychology*, 1933, 4, 156–175.
11. Levenson, S., "The Dialect Comedian Should Vanish," *Commentary*, 1952, 14, 168–170.
12. Middleton, R., and J. Moland, "Humor in Negro and White Subcultures: A Study of Jokes among University Students," *American Sociological Review*, 1959, 24, 61–69.
13. Murray, H. A., Jr., "The Psychology of Humor: II. Mirth Responses to Disparagement Jokes as a Manifestation of an Aggressive Disposition," *Journal of Abnormal and Social Psychology*, 1934, 29, 66–81.
14. Omwake, L., "Factors Influencing the Sense of Humor," *Journal of Social Psychology*, 1939, 10, 95–104.

15. Rosenberg, B., and G. Shapiro, "Marginality and Jewish Humor," *Midstream*, 1958, 4, 70–80.
16. Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill, 1956.
17. Smith, C. U., and J. W. Prothro, "Ethnic Differences in Authoritarian Personality," *Social Forces*, 1957, 35, 334–338.
18. Smith, N. V., and W. E. Vinacke, "Reactions to Humorous Stimuli of Different Generations of Japanese, Chinese, and Caucasians in Hawaii," *Journal of Social Psychology*, 1951, 34, 69–96.
19. Stump, N. F., "Sense of Humor and Its Relationship to Personality, Scholastic Aptitude, Emotional Maturity, Height, and Weight," *Journal of General Psychology*, 1939, 20, 25–32.
20. Wolff, H. A., C. E. Smith, and H. A. Murray, "The Psychology of Humor: I. A Study of Responses to Race-Disparagement Jokes," *Journal of Abnormal and Social Psychology*, 1934, 28, 341–365.